1-1-2003

# Advancing the counterfactual analysis of causation.

Ethan R. Colton
*University of Massachusetts Amherst*

# ADVANCING THE COUNTERFACTUAL ANALYSIS OF CAUSATION

A Dissertation Presented

by

ETHAN R. COLTON

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2003

Philosophy

ADVANCING THE COUNTERFACTUAL ANALYSIS OF CAUSATION

A Dissertation Presented

by

ETHAN R. COLTON

Approved as to style and content by:

Phillip Bricker, Chair

LeRoy F. Cook, Member

Edmund L. Gettier, III, Member

Jonathan Schaffer, Member

Phillip Bricker, Department Head
Philosophy

# DEDICATION

To family,

past and present,

all patient and supportive.

ACKNOWLEDGEMENTS

I would like to thank my committee members for their efforts, comments and suggestions at various points along the way. Thanks especially to Phillip Bricker for his thoughtful criticisms and for his guidance through much of the reading material behind this dissertation.

ABSTRACT

ADVANCING THE COUNTERFACTUAL ANALYSIS OF CAUSATION

FEBRUARY 2003

ETHAN R. COLTON, B.A., BOWDOIN COLLEGE

M.A., TUFTS UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Phillip Bricker

What does it mean to say that one event is a cause of another? The simplest counterfactual analyses identify causation with one of two counterfactual-dependence relations: (1) if event *c* had not occurred, then (distinct) event *e* would not have occurred; (2) if *c* had not occurred, *e*'s probability would have been lower. These analyses enjoy some success. For the first: the dart-throw caused the balloon-pop, because if the throw had not occurred, the pop would not have occurred. For the second: suppose two radioactive samples, A and B, are introduced into a room containing a Geiger counter, and the counter clicks once due to an emission from an A-atom; then the introduction of A is a cause of the click; the click might have occurred without the A-introduction (a B-atom might have emitted), but we can say at least that without the A-introduction, the *probability* of the click would have been lower.

Ultimately, however, these analyses fail, for two clear reasons. *Preemption*: add to the dart scene that Lucy would have thrown her dart if I had refrained—then although my throw caused the pop, the pop is not dependent on my throw. *Failed potential causes*:

the probability of the click would have been lower without the B-introduction, but the B-introduction did not actually succeed in causing of the click.

I defend the two simple analyses against various other objections; I then try to home in on the precise nature of their genuine problems; I examine almost all the attempts to date to improve upon the simple analyses; and finally I propose two new analyses. One of my analyses is deterministic, the other is confined to worlds that are purely indeterministic. Both analyses take causation to be primarily a matter of counterfactual dependence *in the circumstances*: holding certain features of the world fixed, effects and their probabilities do indeed depend (almost exclusively) on their causes (or their direct causes).

My discussion of background issues—counterfactual semantics, objective chance, events—includes arguments for substantial simplifications of David Lewis's theory of events.

CONTENTS

LIST OF FIGURES

# GLOSSARY

The special terminology one might need to reference is defined here. It is also defined in the main text, and the definitions are there attributed, when necessary. In this glossary, lower-case variables range over events unless otherwise indicated.

*backtracking counterfactual*: a counterfactual conditional that says how the past would have been different had the present been different

*b* is *causally dependent* on *a*: *b* is *counterfactually dependent* on *a*, and *a* and *b* are distinct

*b* is *counterfactually dependent* on *a*: were *a* to occur, *b* would occur, and were *a* not to occur, *b* would not occur

*Dependence Thesis*: if *occurrent e* is *counterfactually dependent* on distinct, occurrent *c*, then *c* is a cause of *e*

*Deterministic Simple Analysis (DSA)*: *c* is a cause of *e* if, and only if, they occur, they are distinct, and *e* is *counterfactually dependent* on *c*

*Event proposition*: any result of Boolean complications of *simple event propositions*

*Failure Thesis*: all clear cases of preemption involve a backup sequence that, thanks to the preempting cause, does not completely come off

*e* is *flexible*: it is possible (logically, metaphysically) that *e* occurs differently (with respect to time, place, manner)

P *iff* Q: P if, and only if, Q

*occurrent* event: an event that actually occurs; same as "actual event"

*a* is a *part* of *b* at world *w*: *a* and *b* occur at *w* and none of *a*'s region is outside *b*'s

*Probabilistic Simple Analysis (PSA)*: *c* is a cause of *e* if, and only if, they occur, they are distinct, and *e* is *probabilistically dependent* on *c*

*b* is *probabilistically dependent* on *a*: for some range *R* of chances, (i) were *a* to occur, the chance of *b* at (the end of) *a*'s time would be within *R*, and (ii) were *a* not to occur, the chance of *b* at that same time would be below *R*

*simple event proposition*: any proposition that a certain event occurs or that a certain event does not occur

*e* is *strict* iff *e* is in*flexible* in every respect (time, place, manner)

# CHAPTER 1

# INTRODUCTION

## 1.1 The Question

It seems obvious that some events cause other events: Lincoln's death was caused by a shooting, the chirping of my clock causes me to wake up, storms cause floods. It is a trivial fact that throughout our personal and professional lives, we assume events have causes and effects, and we devote often considerable energy to finding out what they are, were, or will be. But what is this "causation"? How could we answer a skeptic who, despite being very observant, still doubts there is causation? He says, "I see shootings followed by deaths, alarms followed by wakings, storms followed by floods, and I observe the detailed patterns of events in and around these...yet I see nothing more—what is the 'causation' you claim exists?"

It may be tempting to identify causation with something the skeptic observes: perhaps wherever there is causation in the actual world there is some special physical relation—something like, for example, a "flow of energy" from cause to effect (Fair 1979). One apparent problem with this answer is that some causal structures do not involve anything like a flow of energy from cause to effect: cutting the brake cable on someone's car can be among the causes of a car crash; but the cutting just removes a potential preventer of the crash, it does not cause the crash by transferring energy to it (Chapter 3). This particular causal structure aside, we can, it seems, imagine causation lacking any special physical relation that might plausibly hold in actual cases of causation: when we imagine that a magic spell is cast or a pin is stuck in a voodoo doll, it does not seem we are positing anything like a flow of energy from cause to effect—on the contrary, the

causation is magical precisely because of the lack of any such physical connection. The skeptic may ask, "What do all these cases of causation have in common such that you want to count them cases of *causation*? Again, I see the events, but what is the 'causation'?" He is asking us what we *mean* when we say one event was a cause of another, such that all and only cases of event causation, be they actual or merely possible, fit the answer. This is the question I am pursuing. It is the search for a broadly applicable definition, or "analysis," of causation.

Some philosophers (e.g., Nancy Cartwright (1979)) ask what it means to say that a certain kind of thing causes another kind of thing—e.g., smoking causes heart disease, falls cause injuries. I am pursuing only the question of what it means to say that a *particular event* caused, or was a cause of, another particular event—e.g., Jim's fall on Tuesday caused the tearing of that ligament.

And some philosophers (e.g., D. H. Mellor (1995)) ask what it means to say that one fact caused another fact. My concern in this dissertation is only with events.

Sometimes we say that the open flame was *a cause* of the fire, or *caused* the fire, or was *the cause* of the fire. I am pursuing analyses that are fit for the first two locutions but not the third (and *best* fit for the first, perhaps). Obviously, events tend to have more causes than just one, contrary to what is suggested by "the cause." In practice, we choose one cause among many as "the cause." What we choose can shift, depending on our epistemic attitudes: Suppose an open flame is very close to some properly installed carpet, the carpet catches fire and the house burns down. Someone who takes carpeting in a house to be normal and appropriate but open flames abnormal and inappropriate (most of us) will likely count the open flame as the cause of the house fire and reject the

presence of carpeting as the cause; but one who takes carpeting in a house to be abnormal and inappropriate but open flames normal and appropriate will likely count the presence of carpeting as the cause. Since what is the cause can shift, what is the cause in a given case must be relative to something that determines the choice. *What*, I do not know, but just for the sake of illustration, here is a very rough proposal: the flame was the cause relative to a "background circumstance" that includes the carpeting but excludes the flame, and the carpeting was the cause relative to a "background circumstance" that includes the flame but excludes the carpeting. It would be interesting to pin down how "the cause" gets chosen, but it is not my aim here; I am pursuing an inclusive causal relation, one that will readily count both the flame and the presence of carpeting as causes of the house fire.

### 1.2 The Counterfactual Approach

A counterfactual conditional is a sentence about how things would, or might, be if such-and-such were the case. Its content is best expressed in the subjunctive: "If P were so, Q would be," or "If P were so, Q might be"—where P (the "antecedent") is the such-and-such and Q (the "consequent") is how things would (might) be. For example, "If Bush were not President, then Gore might be President"; "Had I won the raffle, I would have thrown a party." Statements of this sort, when P is about the past or the present, are typically made with the presumption that P is not *actually* so. The speaker wants the audience to consider something contrary to actual fact—hence the name "counterfactual." Nevertheless, of course, in a given case it may be that P *is* true, whether the speaker knows it or not; and if it is, the conditional may still be sensible and true or false. "Were it the case that Tolstoy wrote a novel called *War and Peace*, then it would be that others

have read it"—that is true, as is its antecedent. And if in fact I *did* win the aforementioned raffle, then the claim that I would have thrown a party had I won is true or false depending on whether I threw a party. With respect to a *future* P, there is no typical presumption by the speaker that P is false; rather, she may just take it to be somewhat open whether P is true ("If I were to bring dessert, he might be insulted"; "If you were to resign, another would step in")—and indeed P may be true. So, calling these conditionals "counterfactual" is a bit misleading. But it is customary among philosophers, so I will stick with it. I will also follow the practice of calling them "counterfactuals," for short.

The counterfactual approach to analyzing event causation is to analyze it primarily in terms of counterfactuals. The starting intuition is that to say event *c* was a cause of event *e* is just to say that *e* would not have occurred without *c*: more carefully, *c* is a cause of *e* if, and only if, had it been that *c* did not occur, it would have been that *e* did not occur (for distinct, occurrent events *c* and *e*).[1] If the match-strike hadn't occurred, the fire wouldn't have occurred; if the lightning hadn't occurred, the thunder wouldn't have occurred; if the match strike hadn't occurred, the thunder still would have (the analysis rightly says that the match strike was *not* a cause of the thunder). For many simple and common cases of causation, or a lack of causation, this simple analysis holds up fine.

---

[1] By "distinct" in this context writers tend to mean the events *necessarily* do not overlap—i.e., they are *essentially* distinct. We can stay agnostic on this, though for clarity I will settle on *essential* distinctness by the time we reach Chapter 3. An *occurrent event* is an event that actually occurs. It is also known as an "actual event." My preference for "occurrent" stems from my adopted ontology of events (described in Chapter 2), according to which events are properties and the occurrence of an event is the instantiation of a property. Many philosophers allow that properties actually exist whether or not they are actually instantiated; in this case, "actual property" need not mean instantiated property. But I want "actual event" to mean occurrent event. It is most explicit just to *say* "occurrent event."

Throughout the dissertation, I use lower-case italic letters for event variables but also for event names. It should be clear from the context whether a given letter is a name or a variable.

This point will be pressed in Chapter 3 (though ultimately we will see that this simple analysis is inadequate).

One of the things that makes the approach attractive is how successful this simple analysis is. But what recommends it most, I think, is the plausibility of the view that the analysans is, without exception, sufficient for causation: if event *e* would not have occurred without event *c*, then *c* is a cause of *e* (for distinct, occurrent *c* and *e*). Counterfactual analyses of causation almost always accept this Dependence Thesis; I think of it as the cornerstone of the counterfactual approach. I subscribe to it myself. (I defend it at points in Chapter 4.)

This is a good place to introduce a central definition from David Lewis (1973b, pp. 164–6): Event *b* is *counterfactually dependent* on event *a* if, and only if, (i) were *a* to occur, *b* would occur, and (ii) were *a* not to occur, *b* would not occur. I think (with Lewis) that for any events *a* and *b* that occur, it follows that (i) holds, and the question of their dependence thus reduces to whether (ii) holds. Given this, the simple analysis says that, among distinct, occurrent events, causation is counterfactual dependence. The Dependence Thesis is that counterfactual dependence between distinct, occurrent events is sufficient for causation.

The counterfactual approach has accelerated in popularity recently, beginning with the analyses of Ardon Lyon (1967) and especially Lewis (1973b). This appears to be the result of the growing sense that counterfactuals can be well understood, can be true or false, and can be given precise formal semantic analysis. We will get a taste of this in Chapter 2.

Could the above simple analysis satisfy my skeptic? "How," he asks, "do you know that the flood would not have occurred without the storm? I didn't observe *that*. Or did something happen at the scene that implies it?" Nothing happened *there* that alone implied the counterfactual. We did not observe anything the skeptic didn't. But we have a general knowledge of how the world works—of the laws of nature—and we know that without the storm, things would have evolved in such a way that the rivers and streams would have been at normal levels and there would have been no flood. It is the combination of the situation supposed and the laws that govern that general region that determines what counterfactuals are true (Lewis 1973a). Wherever, if anywhere, the laws "are," they are not fully within the causal scene in question. The observant skeptic just failed to pull in his knowledge of the ways of the world.[2]

Naturally the skeptic wants some information about how the laws and the relevant situation imply counterfactuals. And he may complain now that the mystery of causation has just become the mystery of laws: "I see no laws, either!" As to the first, some information is provided in Chapter 2. As to the second, it should be pointed out that the question of what are laws of nature is one that exists anyway, so that if we can reduce the question of what is causation to the (very difficult) question of what laws are, progress has been made by way of a reduction of questions.

---

[2] I am, of course, not talking about "causal laws," laws of the form "every F causes a G." If the counterfactual were true in virtue only of causal laws, the counterfactual approach to causation would be circular. I take it that causal laws, like causal facts and counterfactuals, are true in virtue of the nearby pattern of events and the non-causal laws. (I try to remain otherwise neutral on the nature of non-causal laws.)

There are many approaches to analyzing event causation. (I shall not survey them here.[3]) I need not claim that the counterfactual approach is the best. I find it promising for the reasons mentioned above. Other approaches are also promising.

## 1.3 The Chapters in Brief

Chapter 2 explains some important background assumptions that I make regarding counterfactuals, events, and the objective chances of events; on the subject of events, I argue for a specific simplification of Lewis's (1986b) theory of events. Chapter 3 examines apparent problems for the counterfactual approach, with the goal of homing in on the definite ones, the ones that most need solving. Chapter 4 evaluates contemporary analyses, with a focus on how well they address those definite problems.

Suppose an atom is bombarded with high-energy particles and the chance of its nucleus emitting a particle within the next second thereby increases from .1 to .9; and then the emission occurs. It seems the bombardment should count as a cause of the emission. This is an example of indeterministic causation: the effect is never determined to occur (its chance is never 1). I am interested in both deterministic and indeterministic causation in this dissertation.

---

[3] Here is a snapshot of the landscape: There are many views (e.g., J. L. Mackie 1965) that take causation to be primarily a matter of instantiating a lawful regularity (some of those views are probabilistic, a matter of one event being, by law, more probable thanks to the other); this is made precise in various ways. Some views take causation to be primarily the instantiation of one of a special class of laws (e.g., Armstrong 1997). Some views are agent-centered, so that a cause is something some idealized agent could use to bring about the effect, or make it more probable (e.g., Menzies and Price 1993). Some define causation in terms of some basic causal relation identified with an actual physical relation and disregard non-actual possible worlds (e.g., Fair 1979). Some take causation to be a theoretical concept, in the sense (Lewis 1970) that the causal relation is by definition *whatever* relation plays a certain (non-causal) role, this player being one thing in some worlds but perhaps different things in other worlds (e.g., Tooley 1987). Often these views contain many elements, including some counterfactual conditionals, spatiotemporal constraints or epistemic components. (It should also be noted that there are those who have thought there is no such thing as causation (e.g., Russell 1912-13), or that causation is primarily an epistemic phenomenon (Hume

To date, counterfactual analyses divide into those restricted to deterministic worlds and those meant to handle both deterministic and indeterministic worlds. I will be examining both the restricted and the broad analyses. A source of this division seems to be the failure of the above simple analysis to handle simple cases of indeterministic causation: notice that if the bombardment hadn't occurred, the emission might have occurred anyway. As we will see in Chapter 3, the broad analyses usually take a different intuition as their starting point.

In Chapter 5 I propose my own deterministic counterfactual analysis. I do not have a broad analysis to offer. However, in Chapter 6 I offer an analysis suited to worlds that are "purely indeterministic," meaning that, at any time, any possible future pattern of events has some non-zero chance of taking place. Our own, actual world appears to be neither deterministic nor purely indeterministic: quantum theory implies that it is not deterministic, while there are still plenty of deterministic constraints in physical theory (such as that nothing can accelerate beyond the speed of light—the chance of something doing so is always 0). Then why do we care about analyses restricted to the extremes of determinism or pure indeterminism? To the extent such analyses are successful, they may suggest, or be useable to create, a broad analysis. Second, the extreme analyses are of some interest directly, since they may analyze two ways we habitually think about the world—sometimes as deterministic, sometimes as purely indeterministic. The narrowed concepts they define may therefore comprise important portions of our concept of causation. (Despite the great attention that deterministic analyses have received in

---

1739, on some interpretations); and there are many who think that counterfactuals should be analyzed in terms of causation rather than the other way around (e.g., Kvart 1986).)

philosophy, an analysis for purely indeterministic worlds is, I think, in line with the more common way people naturally or naively think about the world: no occurrence or nonoccurrence really has *no* chance; some are just *very* improbable, to the point of being freaky or "miraculous.")

Still, my purely indeterministic analysis is itself something of a draft. For it is, I believe, unsatisfactory in at least two respects, and I am not presently able to improve upon it. First, it is inelegant—inelegance, if irreparable, suggests to me falsity. Second, as I explain in Chapter 6, certain components are not fully motivated.

# CHAPTER 2

# IMPORTANT BACKGROUND ASSUMPTIONS

## 2.1 Introduction

The analyses discussed in this dissertation employ counterfactuals (Section 1.2) and make reference to events and chances of events. Throughout, I rely upon certain assumptions about these three things that I shall now describe and explain.

## 2.2 The Interpretation of Counterfactuals

The analyses I discuss throughout the dissertation generally assume David Lewis's views about counterfactuals, and I shall do so as well. Here I describe the important components.[1]

It is intuitively plausible that when I say, "Were I to flip the switch, the lamp would go on," I am saying something like this: among possible worlds where I flip that switch, the worlds akin to ours are worlds in which the lamp then goes on. Notice the loss of the subjunctive mood; the intuition reduces the counterfactual to a certain similarity comparison between kinds of possible world. Lewis (1973a) developed a semantics along these lines, according to which a counterfactual "If P were so, Q would be" is actually true iff (if, and only if) either (i) P is impossible or (ii) there is a possible world at which P and Q are true and which is closer to—that is, more similar to—the actual

[1] Throughout the dissertation, I assume a possible-worlds framework. A necessary truth (e.g., 2+2=4) is true at all possible worlds; an impossibility, or necessary falsehood, is false at all possible worlds. Other sentences, statements, propositions, are true at some worlds but not others; to be actually true is to be true at the actual world. For example, the proposition *that there are diamonds* is true at the actual world; but it could have been false, which means that it is false at other possible worlds lacking diamonds. (A "proposition" P is often identified with the class of possible worlds at which P is true.) As some others do, I take "actual world" to be "indexical": like "I," what it designates shifts with contexts and speakers, and it can be used to refer to any possible world.

world than is any world at which P is true but Q false. For instances in which there is a *closest* P-world (a closest world at which P is true), or a set of P-worlds tied for closest, it follows that the counterfactual is true iff the closest P-worlds are all Q-worlds. For simplicity, I will usually assume that if P is possible then there are closest P-worlds (Lewis refers to this as the Limit Assumption (1973a, pp. 19ff)).

The semantics itself does not dictate the nature of the similarity relation relevant to a proper interpretation of a given counterfactual. Lewis allows that in different contexts we may intend different standards of similarity, weighting similarity with respect to some properties more or less. We must look to our judgements in a given context of which counterfactuals we take to be true and which false if we wish to know (roughly) what similarity relation is in play in that context. Consider: "If Caesar were in command, he would use the atom bomb"; "If Caesar were in command, he would use catapults" (Quine 1960, p. 222). The first sounds true when we take a world in which present-day Caesar has adapted to modern warfare to be more similar to ours than a world in which present-day Caesar has kept his ancient techniques; the second sounds true when we judge it the other way. The first apparently invokes a similarity relation that gives a lot of weight to general traits of Caesar such as his opportunism, his desire to utterly destroy an enemy and his preference for the most effective weapons available; the second apparently invokes a similarity relation that gives more weight to Caesar's weapon-specific habits, skills and knowledge.

It seems to some people, me included, that without exception no world is as or more similar to a world *w* than *w* is to itself. Therefore, any counterfactual with true

antecedent is true iff its consequent is true. (This is not a topic of controversy among any authors discussed in this dissertation.)

A *backtracking counterfactual* is a counterfactual that says how the past would have been different had the present been different (Lewis 1979, pp. 32–5). Lewis thinks there is a certain standard use of counterfactuals on which the intended similarity relation makes backtracking counterfactuals false. For example, suppose we know that Charlie was on a high cliff with just his usual shorts, T-shirt and daypack. You say, "If Charlie had jumped off the cliff, he would have gotten badly hurt." But I say, "No, if Charlie had jumped, he would have put on a parachute earlier (because Charlie would never jump without a parachute, he's such a sensible guy). So if he had jumped, he would have landed unhurt." Clearly we are reasoning in different ways. This, according to Lewis, reflects that we are invoking different sorts of similarity relations in our use of the counterfactuals: my first counterfactual assertion invoked a non-standard sort of relation; your counterfactual invoked a standard one. On the standard sort, my assertion was false—had Charlie jumped, the relative past would *not* have been different such that he would have put on a parachute. As Lewis notes, it is often more comfortable to insert the phrase "have to" when backtracking: "If Charlie had jumped, he would have to have put on a parachute." The natural inclination to use this special phrasing supports the idea that backtracking is a special usage.

It is the more standard usage that Lewis intends for the counterfactuals in his own causal analyses. Other analysts, myself included, have followed Lewis in drawing this distinction between backtracking and non-backtracking sorts of similarity relations and in restricting their analyses to the latter. The restriction is important, because it paves the

way for causation to be asymmetric on a counterfactual analysis. As I mentioned in Section 1.2, counterfactual analyses generally imply that counterfactual dependence between distinct, occurrent events is sufficient for causation. Suppose I press the doorbell button and this causes the bell to sound. We might be inclined to accept, "If the sound had not occurred, it would (have to) have been that I did not press the button"—that is, the pressing is counterfactually dependent on the sound. If counterfactual analyses were not restricted to a similarity relation on which such backtracking counterfactuals are false, the quoted counterfactual would imply that the sound was a cause of the pressing.

While Lewis's view that backtracking counterfactuals are in some sense special has been challenged, I will largely leave that debate aside. I am inclined to think that Lewis is correct, at least for the sorts of event-based counterfactuals used in causal analyses; and in light of the various challenges, it would be a major project to defend the view. I shall just adopt it as an assumption.

The *reasons* Lewis is correct (if he is) must not include that we make certain causal presuppositions in our standard interpretation of counterfactuals, else a causal analysis relying on such an interpretation will be circular. But Lewis (1979) attempts to give a general description of the sort of similarity relation we standardly invoke in counterfactuals about events, and it does not rely on causal concepts. I take it as a working assumption that some such non-causal description of the relevant sort of similarity relation is possible, either Lewis's or some other. I said that the restriction to a non-backtracking kind of similarity relation "paves the way" for causal asymmetry; it is this non-causal description of the relevant kind of relation that is needed to bring it home.

Lewis offers two interpretations of the "might" counterfactual.[2] The central one, and the one I will always intend in my use of it, is this: (S1) "If P were so, Q might be" means the same as (S2) "It is not the case that if P were so, Q would not be." S1 is consistent with (S3) "If P were so, Q might not be." So notice that since S3 means the same as (S4) "It is not the case that if P were so, Q would be," S2 is consistent with S4. Also noteworthy is that a "would" counterfactual always implies the corresponding "might": if it is true that were P so, Q would be, then obviously it is false that were P so, Q would *not* be, and hence by definition it is true that were P so, Q might be.

Both a "might" and the corresponding "might-not" may be true when there is significant underdescription by a false antecedent. Thus, if Bizet and Verdi were compatriots, they might have been French—and they might *not* have been French, but rather Italian. (Perhaps also they might have been citizens of some other country—it seems to me that in supposing that they are compatriots, one may or may not intend to invoke a similarity relation whereby the two men might belong to some third country.) To use events: If there were a musician practicing in the next room, we might be hearing bassoon sounds now, and we might be hearing trumpet sounds. Likewise, even in a deterministic setting we should say that if I were to flip this coin, it might come out heads, it might come out tails. Were I *precisely* to specify the nature and circumstances of the flip, *then* I could say truly that it would come out heads, or I could say truly that it would come out tails.

In these examples, I suppose things *were* to obtain that did not. Most counterfactual analyses are only concerned with supposing that events which did occur *were not* to

---

[2] Discussed in Lewis 1973a (pp. 2, 21ff) and 1979 (pp. 63–4).

occur. There is disagreement about how much underdescription this involves. Some think that in supposing away an event we mean to suppose that the event is "completely and cleanly excised from history" (Lewis 2000, p. 190). Others (Michael McDermott (1995a) and Stephen Yablo (2002)), including myself, think that we sometimes allow that parts of the event might (and might not) remain. "If his bicycle trip across the country had not taken place, he might still have at least started it...though he might not have even done that"; "If the Great Chicago Fire hadn't occurred, still one barn fire might have occurred...or it might have been that *all* of the Chicago fire was absent"—these sound fine, to me. (And as we will see in later chapters, such "might" counterfactuals are needed to defend against certain alleged counterexamples to various analyses.)

A second situation in which both a "might" and the corresponding "might-not" can be true is where (the antecedent is false and) the consequent involves an indeterministic event: "If there were an unstable atom on my desk, it might decay in the next five seconds." It is also true to say that the atom might *not* decay in the next five seconds.

Lewis generally presumes, as will I, that if in some non-actual situation there would be a non-zero chance that an event *e* occurs (does not occur), it is true to say that *e* might occur (not occur), where the "might" here is interpreted in the way just described. (He allows exceptions, but they need not concern us.[3])

---

[3] I have in mind Lewis's "convergence quasi-miracles" (1979, pp. 58–64). He calls a pattern "quasi-miraculous" when it results from chancy events "that seem to conspire" to produce it. He imagines that all the traces of the pressing of a button—the signal down the wire, the heat in the insulation, the light images of the pressing moving toward outer space, the fingerprint—have some chance of dissipating such that the world is exactly as it would have been had no button-pressing occurred. Such a cover-up would be a convergence quasi-miracle. We are drawn to say that, despite having some chance (and thus not being a miracle simpliciter), such a thing would not happen were the button to be pressed. Lewis agrees that it would not; its occurrence, he says, would be too "unlike the goings-on we take to be typical of our world" (p. 60) for it to be among any of the closest worlds in which the button is pressed. However, he admits that in such a case there is still a sense in which the quasi-miraculous event might occur, though not the sense I describe in the main text. On this other sense, "it might be that P" means *it would be possible that P*. Since

Now, we do say, "If he were to play the lottery, he wouldn't win," and "If all three women had come to dinner, surely they would not all have worn the same dress." Here we deny that a certain thing might happen even while, it seems, being aware that it would have *some* chance. It seems to me that in these examples, the motivation behind saying that the improbable event would not occur is just how very improbable it is. Should I not, then, allow that, due to an event's very low (but non-zero) chance, it would not occur?

In such statements as the above two, I think we may be suffering from the false sentiment that things need a certain minimum non-zero probability in order to happen—"That's so improbable, it could never happen!" Or, we may be engaged in the practice of setting small chances aside: "The chance of that is negligible"; "The chance of that is so small, I think we can just ignore it." Exactly why and when we engage in this sort of rounding down, I do not know. But counterfactuals that rest on it may fail to be literally true. I think that when a statement that some future event would not occur is based *only* upon the fact that its chance would be very low (but non-zero), the statement is false. Imagine that 1 trillion people could draw straws and that, if they were to do so, exactly one person would get the winning straw. Suppose that, just because of the odds, we are moved to say, "If they were to play that game, #300 would not win—some other player would." We ought, then, to say the same of #1, #2, etc., since there is nothing special about #300. Since no player is special, either all of these trillion counterfactuals are true or none is. Yet they *cannot* all be true, unless they invoke different similarity relations.

---

the event's chance would be non-zero, it would be possible for it to occur: it "might" occur (in one sense)...even though it would not (in another).

If they invoke the same one, then *each one* of the closest worlds in which the people play is, according to the consequents, a world in which #1 does not win, #2 does not win, ... and yet some player wins—impossible. But there is no motivation for *different* similarity relations, since the counterfactuals all have the same antecedent and there are no relevant differences between the different players mentioned in the consequents. There is nothing more similar to actuality about a world in which they play, #1 loses and someone else wins than a world in which they play and #1 wins; likewise for all the other numbers.[4]

## 2.3 Chance

We will see analyses that use the notion of the "chance" of an event—the chance that the event occurs. Chance is single-case, objective probability. It is "single-case" in that it is not a ratio or frequency of any type (so we can ignore worries about what is the appropriate "reference class"). It is "objective" in that it does not hold only relative to some believer or body of information—it is not an epistemic probability of any type. That right now there is a certain chance x that that nucleus will emit an alpha particle within the next five minutes is an objective fact about the world, as believer-independent as any other.

The most important point for us is that chances are standardly time-relative: At $t$, the chance that S is x. The chance of an event may vary through time, as conditions vary. Lewis (1980, p. 91) gives the example of someone working through a maze. The person begins with a moderate chance of reaching the center by noon. As he moves about, half

[4] Imagine you are rationally justified in believing you do not actually play a certain lottery of very long odds. If the reasoning I have just given is correct, it seems it should be rational to believe that if you *were* to play that lottery, you might win. Now, it is often said that if you know you play a lottery of very long odds, then it is rational to believe you lose—the belief is justified by its high probability. There is a curious tension here, but I see no contradiction.

an hour later he has reached a region from which it is much less probable he will reach the center by noon. Despite the odds, by 11:45 he has escaped the difficult area and is in a good position: his chance of reaching the center by noon is now quite high. At 11:49 he has reached the center, so at this time his chance of reaching the center by noon is 1. And it will be 1 at any time in the future. Likewise, in quantum mechanics, the chance that a system has, or will have, a given property can vary through time as the system's state varies. It is generally assumed that the chance of any past event is 1 if it occurred, 0 if it did not, and I follow that assumption.

I assume that the chance of an event at time *t* is determined by the world's entire past event history up through *t* and the (non-causal) laws of nature for the world.[5] Since this history and these laws are contingent (they do not hold logically or necessarily), any particular actual chance is relative to the actual history and laws—that is, relative to the actual world. So throughout I shall assume a single chance function that takes a time, a world, and a proposition that a certain event occurs and yields a real or infinitesimal number in [0,1]. (The emission-from-the-nucleus-*at-that-precise-time* was an event whose earlier chance was infinitesimal; so we ought to allow infinitesimal chances.)

## 2.4 Events

It will pay to assume some fairly precise picture of events, including a theory of the part-whole relation between events. I shall largely adopt Lewis's (1986b) views on these, though I shall deviate a bit.

---

[5] Throughout, I make no special assumptions about (non-causal) laws, except that their expression need not involve any causal concepts: as discussed in the previous chapter, the causal facts exist partly in virtue of the (non-causal) laws, not the other way around.

### 2.4.1 The General Nature of Events

At least for present purposes, every particular event can occur only once. Thus I do not count the Boston Marathon that occurs every year as a particular event, while the different non-repeatable marathons are particular events. I will defend this stance in a moment. For every occurrent event there is a smallest spatiotemporal region in which the whole of it occurs and outside of which none of it occurs—I will refer to this region variously as the region *in* which the event occurs, as the region *of* the event, and as "the event's region."

Each particular occurrent event is a certain property of the region in which it occurs. Lewis takes such a property (event) to be the class of possible regions in which it is instantiated (occurs). An occurrent event is any such class with a member from the actual world. No event has more than one member from any one world (events do not recur, as I said above). No event has a member from every world (no event necessarily occurs). If a class of regions meets these criteria, it is what Lewis calls "formally eligible" to be an event; then there are further constraints, which we will visit in Section 2.4.3.

It seems there may really be repeatable particular events, contrary to what I have said. We say the Boston Marathon occurs every Spring, the eruption of Old Faithful occurs regularly, the Olympics occurs every two years, high tide happens twice a day. Here we talk as if repeatables are particular events. On the other hand, perhaps it is loose talk, and we *really* mean that the Boston Marathon is a *kind* of event, each *member* of which occurs in a Spring—exactly one Spring. We sometimes pick out this year's Boston Marathon by saying "*this particular* Boston Marathon is especially competitive"; we count how many Boston Marathons we have seen, just as we count how many house fires

or weddings we have seen; so it is also consistent with some of our talk that the repeatable is a kind from which we pick out particular individuals, where the individuals are not repeatable. I see nothing to settle whether the repeatable is a particular event or not—and it may even be both: sometimes we mean it one way, sometimes the other.

While there may really be repeatable particular events, I want to ignore them, so I shall just declare that I am not counting them as particular events within this dissertation. Why? It is not because they are not involved in causation; there is *some* evidence that they are: "the Boston Marathon caused a traffic jam this year, but it did not cause one last year"; "high tide is an effect of the moon's position."[6] Repeatables are extra work, because we have to agree on the causal facts about them (was the 1999 starting gun a cause of the repeatable marathon?) and we have to decide what it means to suppose that the repeatable were not to occur (that *none* of it occurs, past and future?). These are not easy questions. I wish to ignore repeatables largely because we have a great deal of work to do as is. It seems the analysis of causation among non-repeatables is a fine standalone first step: as far as I can find, the repeatable Boston Marathon is a cause or effect only in virtue of some of the non-repeatable Marathons that occur a year apart, so that understanding causation among non-repeatables may be a useful first step to understanding causation among repeatables. In ignoring repeatables I will not, by the way, be skirting challenges that others have faced: repeatables have always (with little, if any, comment) been excluded from analyses of particular-event causation.

---

[6] Here I part from Lewis, who rejects (without argument) that repeatables ever "cause or get caused *simpliciter*" (1986b, p. 243). I see no way to insist upon that. (Even if all causation among repeatables can be reduced to causation among non-repeatables, it of course does not follow that there is no causation (*simpliciter*) among repeatables.)

Lewis allows that events may be very fine-grained ("allows" in that his constraints will not prevent it, and he openly embraces it). He gives an example in which John says "Hello" rather loudly—thereby there occurs, in one region, an event that is essentially a hello and essentially loud as well as an event that is essentially a hello but is possibly soft (1986b, p. 255). And given Lewis's general discussion in 1986b, there surely occur in that region even more variants, such as an event that is essentially a verbal greeting but only inessentially a hello, and many others. How liberal we are to be is not pinned down by Lewis. I shall try to pin it down in Section 2.4.3.

Such a generous ontology of events is unusual,[7] and it may seem excessive. I do make use of it, so let us see how it may be justified.

Lewis argues that we cannot get by with just a hello that is essentially loud, nor just with one that is not, because each plays a causal role, and they do not play the same role (p. 255). The essentially loud hello is an effect of John's state of tension (an event)—it would not have occurred without that tension—but it is *not* a cause of Fred's return hello—the return hello might still have occurred even if the essentially *loud* hello had not (because John might have said "Hello" at a soft volume). On the other hand, the hello that is only *in*essentially loud is *not* an effect of the tension—it would still have occurred if the tension had not (it would have occurred as a *soft* hello)—and it *is* a cause of Fred's return hello, since the return hello would not have occurred without it (without it, there would have been no hello at all, loud or soft).

Lewis's causal claims here may appear to rely on the associated counterfactual dependences, and one might doubt these or question whether they are decisive evidence

[7] Not unique, however. Kim (1973b, 1976) also allows that ordinarily multiple events occur in one region.

of causation in this story. So let me further champion the plurality of hellos without relying on these subtle causal and dependence claims, by drawing attention to some work that they do.

Suppose it is true that when John says "Hello," only one hello occurs. Let me embellish the story. This hello was John's greeting when he came in to work, and he always greets his officemates with a hello. Officemates Fred and Fran always respond to this greeting by saying "Hi." This morning, when John, tense, says "Hello" rather loudly, Fred says his usual "Hi," and Fran says "Hi," too, though she also thinks to herself, "Gee, that was rather loud...I wonder if he's alright." Clearly, John's state of tension was a cause of Fran's thought; and this causation was obviously not direct, but mediated by John's utterance, a hello—*the* hello, allegedly, since we are supposing there is only one. And clearly, a hello by John was a cause of Fred's return greeting. But if there is only one hello in this story, then John's tension caused the hello that caused Fred's return greeting. If causation is transitive,[8] we get that John's tension was a cause of Fred's return greeting—surely that's not right. Add that Fred wears a hearing aid that cannot detect the additional volume in John's voice; then it is even more counterintuitive that John's tension was a cause of Fred's return greeting. The only way to deny that the example refutes transitivity while denying that John's tension is a cause of Fred's return greeting is to accept the remaining alternative, that two hellos occurred. Which should we do?

On the one hand, the example does appear to refute transitivity. On the other hand, something sneaky seems to be going on, and we should not sweep it under the rug. It is

[8] Event causation is transitive iff whenever an event x causes an event y and y causes an event z, x causes z.

alleged that there is a causal chain from John's tension to the hello to Fred's greeting. But our sense is that there is slippage at the middle link: what the tension is itself responsible for is not anything that in turn is responsible for Fred's greeting. The alleged chain sits ill with that sense. We want to say something like this: the tension caused the *loudness* of the hello, but this in turn is not what caused Fred's greeting. These intuitions should make us reluctant to reject transitivity just on the basis of the example. On the theory that there are at least two hellos here—one essentially loud, one inessentially loud—we can respect and make sense of these intuitions. Suppose we follow Lewis, above, and accept that the tension is only a cause of an essentially loud hello, which hello in turn is not a cause of Fred's return greeting, and the tension is *not* a cause of a hello that is *in*essentially loud, though this hello *is* a cause of Fred's return greeting. Then the discomfitting causal chain is gone; the suspicion of slipperiness at the middle link is vindicated and explained; the intuition that the tension just caused the loudness of the hello is cashed out ontologically; and the preference against rejecting transitivity on the basis of the example is satisfied. That is a good amount of work accomplished.

### 2.4.2 Spatiotemporal Mereology of Events

I will have occasion to talk of event composites and event parts, and it will pay to be clear about what I mean. Events are, I have said, classes. Perhaps the clearest way to understand "part of a class" is that it means *subclass*.[9] But, as Lewis says (1986b, pp. 258), when talking of event parts, we may—and I always shall—have something else in mind, something spatiotemporal: a big event's region has subregions that are the regions

[9] Lewis takes subclasses to be parts of classes (1986b, p. 258; 1991)

of parts of that big event; and events in two regions can make up a whole that is an event that occurs in the union of those regions. I need what Lewis refers to as a "spatiotemporal part" relation for events. I will give a definition of this relation (and of derivative relations, such as "overlap"). Then I will describe some general truths about it. A *mereology* is a theory about the part-whole relation; the general truths I describe can be taken as axioms in a "spatiotemporal mereology" of events (Lewis 1986b, pp. 258–60).

An important fact about (spatiotemporal) parthood of events that needs to be respected in this project is that some events could occur differently with respect to their parts—with more parts, or with fewer parts, or with different parts: that concert, we usually think, might have occurred with an additional encore, with no encore, with a different encore, etc. So parthood needs to be world-relative: at the actual world, the concert had one Beethoven encore, but at other possible worlds it had two of them, at others none, at others Paganini. The "part of" relation (and other related notions) will be three-place, where a world occupies one place: $e$ is a part of $f$ at $w$.

I shall use simpler definitions than Lewis's. As I explain in footnote 10, I believe the added complexity of Lewis's definitions fails to achieve what he aims for it to achieve. (Though if you wish, you are free to assume Lewis's definitions—given in footnote 10—instead. This will not compromise any of my important arguments in later chapters.) I follow, by the way, the common practice of allowing that anything is a part of itself—an "improper" part, as opposed to a proper part. Likewise, I take a spatiotemporal region to be an (improper) subregion of itself. Here are the definitions: event $e$ is a *(spatiotemporal) part* of event $f$ at world $w$ iff $e$'s region at $w$ is a subregion of $f$'s region at $w$; events *overlap* at $w$ iff some event is a part of each at $w$; $e$ is a *sum* of events $f_1, f_2$,

... at $w$ iff $e$ overlaps at $w$ all and only those events that overlap at least one of the $f$'s at $w$; events $f_1, f_2, \ldots$ are *distinct* at $w$ iff no event is a part of any two of them at $w$.[10]

As a point of comparison, let us note three common axioms of mereology:[11] (Transitivity) if x is a part of y and y is a part of z, then x is a part of z; (Unrestricted Composition) for any things, there is at least one sum of those things; (Uniqueness of Composition) for any things, there is at most one sum of those things. We would need world-relative versions of any of these we use. It will turn out that the axioms we need are world-relative versions of two of them.

---

[10] Here are Lewis's definitions. That of "part" proceeds in stages: *e implies f* iff, necessarily, if *e* occurs in a region then also *f* occurs in that region; *e* is *essentially part* of *f* iff, necessarily, if *f* occurs in a region, then also *e* occurs in a subregion included in that region (not necessarily a proper subregion); occurrent event *e* is *part* of occurrent event *f* iff some occurrent event that implies *e* is essentially part of some occurrent event that implies *f*. (In terms of event classes, occurrent event *e* is part of occurrent event *f* iff there are event subclasses of *e* and *f*—$e_s$ and $f_s$—such that every member of $f_s$ has a subregion that is a member of $e_s$, and $f_s$ has an actual member.) "Events overlap iff they have some event as a common part; [...] an event *e* is the mereological sum of events $f_1, f_2, \ldots$ iff *e* overlaps all and only those events that overlap at least one of the *f*'s." "Distinct" is never explicitly defined. (By the way, outside the context of this discussion, I shall not mean by "implies" what Lewis means. By "*e* implies *f*" I will only mean that necessarily, if *e* occurs, then *f* occurs.)

Parthood often holds contingently; an occurrent event is part of another occurrent event *at the world of occurrence*. Without the indexicality of "occurrent," Lewis's "part" definition comes to this: event *e* is part of event *f* at world *w* iff some *e*-implying event that occurs at *w* is essentially part of some *f*-implying event that occurs at *w*. Since the remaining mereological notions are defined in terms of part, the whole mereology is as world-relative as the one I give above.

Lewis uses this definition of "part" because he thinks events that occur in the same actual region may be actually distinct. One example he gives is of a philosophy conference and a battle of goblins composed of matter that does not interact with ours. The idea is that, in such cases, it won't be true that the two events or any occurrent events that imply them stand as essential part and whole; where co-occupants are in fact *not* distinct, there will at least be occurrent events with very narrow, strict essences that imply them and stand as essential part and whole.

For that to work, it has to be that two distinct co-occupants never have a sum that implies both of them; any such sum is, of course, an essential part of itself, so the co-occupants would come out actual parts of each other on Lewis's definitions. An example of such a sum would be an event that essentially occurs in the region the two co-occupants occurred in and that occurs only if the two occur in that region. I see no principled way to rule out such sums. I would go further and say that there are *always* such sums, since, as I explain in the main text, I accept an unrestricted composition principle; this would mean that Lewis's definition just reduces to the simple one I am using. That is, I accept the following principle: for *any* occurrent events *a* and *b* that occur in a region R, there is an occurrent event *g* that essentially occurs in R and occurs only if *a* and *b* occur in R. Again, since *g* is an essential part of *g*, *a* comes out a part of *b* on Lewis's definition. Thanks to that famous Australian battle-conference of 1981 (*g*), the battle (*a*) and the conference (*b*) are parts of each other after all.

[11] Taken from Lewis (1991, p. 74); he refers to these as the basic axioms of mereology.

Spatiotemporal regions have a mereology that is transitive. A part of a region is a subregion; if region x is a subregion of region y, which in turn is a subregion of region z, then x is a subregion of z. This fact, in conjunction with the definition of "part," forces upon us our first axiom: (World-relative Transitivity) If *a* is a part of *b* at *w* and *b* is a part of *c* at *w*, then *a* is a part of *c* at *w*. I assume this second axiom: (Unrestricted World-relative Composition) For any events $f_1, f_2, \ldots$ that occur at *w*, there is at least one event that is a sum of the *f*'s at *w*. Lewis is explicitly undecided as to whether to accept this axiom, because, clearly, it generates some peculiar sums (1986b, p. 260). But I can find no good reason to think there are no such events, peculiar as they are. I will defend this stance in the next section. Those are the only axioms of the mereology. We will see just below why there is no uniqueness axiom.

The claims that not all events have their parts essentially and that different events, like the different hellos, regularly occur in the same region, are substantive metaphysical claims about events that, independently of the two mereological axioms, generate two noteworthy facts about sums. First, since not all events have their parts essentially (the concert could have had a different encore) it follows that an event may be the sum of certain events at one world but the sum of other events at another world: the concert is actually a sum of the main program and the Beethoven encore, but possibly it is a sum of the main program and a Paganini encore.

The second fact is that events have multiple sums at one world. Recall the various occurrent hellos—some essentially loud, some that could have been soft, some that, while essentially greetings, could have been howdys instead of hellos (Section 2.4.1). They occur in the same region, so they actually have all the same parts (including each other,

as non-proper parts);[12] but then the parts have multiple sums at the actual world, namely the various hellos themselves. An axiom of uniqueness of event composition would say that for any events $f_1, f_2, \ldots$ that occur at $w$ there is at most one event that is their sum at $w$. But we see now that we cannot adopt this axiom.

Because of the many sums at one world, we have to be a little careful about talking of *the* sum of the events. Of course, we *can* talk about *the* sum, on the presumption we know at least roughly which one we are talking about. That is, we can make the sort of presumption we generally do: when we talk about the dinner party or yesterday's storm, a variety of events, with varying occurrence conditions,[13] could be intended, but we ordinarily presume we have at least vaguely the same one in mind.

Another substantive thesis should be mentioned. Multiple events occur in one region not only due to differences in essentiality of intrinsic features such as loudness, but also due to differences in essentiality of spatiotemporal proper parts. Just as we may talk about both a hello that is essentially loud and one that is not, in one region, also we may talk about a concert that essentially has an encore as a part and one that does not, in one region.[14] How liberal we are to be about what sorts of variation in part essentiality there

---

[12] To see this on Lewis's "part" definition, notice that an occurrent hello with a very narrow, strict essence will imply all the occurrent hellos and their parts and is essentially part of itself.

[13] An event's "occurrence conditions" are what is necessary and what is sufficient for the event's occurrence.

[14] The motivations carry over. Imagine that John is the conductor. The applause at the end of the main program is about to die out—the audience does not want an encore—but an encore occurs anyway because John is nervous and fails to wait as long as he should for the applause to die out before beginning one. Later that night, Fred reports to his mother that he saw John's first concert—something he would have reported whether there had been an encore or not. Was the nervousness a cause of a concert and this very concert a cause of Fred's report? Was the nervousness a cause of the report? No, and No. The nervousness was a cause of a concert that has an encore essentially, though not of one that has it inessentially; and only a concert that *in*essentially has an encore is a cause of Fred's report. The relevant counterfactuals support this; and it does the work of allowing us to deny the alleged causal chain that seems a sneaky attack on transitivity.

may be is not pinned down by Lewis, but I shall pin it down in the next section. (Notice that just as the multiple hellos in one region gave us multiple sums at one world, the multiple concerts do, too; their parts, such as the main program and the encore, have as a sum the concert that has the encore essentially and also the concert that has it inessentially.)

### 2.4.3 Restrictions on "Formally Eligible" Event Classes

The classes that are "formally eligible" (Section 2.4.1) to be events are the non-empty classes of possible spatiotemporal regions, no class with two members from the same world, no class with a member from every world. Lewis never completely defines what an event is, but he adds three eligibility requirements to the formally eligible classes. I will argue that the second of these (in the order of my discussion) follows from the first and that the third is unnecessary—really we need only the first. I shall also, as promised, defend the mereological axiom of unrestricted composition. And I shall answer the question of what sorts of variation with respect to essentiality of features and parts there may be—I do this by giving a rough definition of "event."

The *first* requirement is that events be "predominantly intrinsic" properties (1986b, pp. 262ff). Events are not *entirely* intrinsic properties, because we want to differentiate intrinsically similar events in different regions—we do not want to count them as the same one event. Lewis gives as an example of an overly *ex*trinsic "event" *the widowing of Xanthippe*, which he defines as something that essentially occurs in region R and occurs iff Xanthippe exists in R and her husband dies at that time (outside R). It is an event whose occurrence is largely a matter of what goes on outside its own region.

That is an odd event—I doubt we would ordinarily say that the widowing of Xanthippe occurred where Xanthippe was sitting. If we knew where to place it at all, I think we would be more apt to say it happened in a prison cell, where the death occurred. There are, in fact, many ordinary examples of very extrinsic events with clear locations: an adultery (an act of committing adultery), a presidential assassination, a parking violation, the breaking of a promise—we at least sometimes talk of these events with the intention that, necessarily, they occur only if certain kinds of events outside their regions also occur. For instance, we may take the adultery to require the presence somewhere of a spouse, while also believing that it occurred in the bedroom (a region excluding the spouse).[15] We may take the presidential assassination, which occurred in the theater, to be something that could not occur but for the victim's being a president and thus having engaged in a certain swearing-in ceremony. We may take the parking violation that happened on that street corner to be an event that could not occur but for the passage of a relevant law. And we may take a certain promise-breaking to be an event that occurs partly in virtue of an earlier promise. Who is prepared to deny that these adulteries, assassinations, parking violations and promise-breakings have their causes and effects? Not me. And do they have their causes and effects *only* derivatively, by way of the causal roles of more intrinsic versions—for instance, might the adultery be an effect *only* by way of the bedroom activity itself being an effect? Not clearly so: The distant husband promised he would sign the divorce papers by midnight; but unbeknownst to her, he changed his mind at the last moment; she, not unreasonably, thinks his reversal is a cause of her engaging in an act of adultery: the adultery is an effect, but not by way of

[15] Like many others, I take the presence of an object in a region to be an event.

the bedroom activity itself, since the latter is obviously not an effect of the husband's reversal. How, then, can excluding very extrinsic events from an analysis of event causation be justified?

Lewis assumes what in Chapter 1 I called the Dependence Thesis: counterfactual dependence between distinct events is sufficient for causation. And he observes that when the Dependence Thesis is applied to very extrinsic events, we get bad results. For example, if Socrates' death had not occurred, the widowing of Xanthippe would not have occurred—yet we do not want to count the death a *cause* of the widowing. The death, as Lewis points up, is just an event *in virtue of which* the widowing occurs, not a cause; it simply satisfies an essential criterion for the widowing. We see the same thing in the case of the adultery. If the simultaneous presence of Husband had not occurred, the adultery would not have; yet the mere presence of Husband seems not like a cause of the adultery, only a satisfier of a necessary condition for the adultery. Lewis uses such bad results to justify the claim that very extrinsic "events" are not really events. Let me explain the reasoning behind that move.

Lewis holds that there may be various sorts of event, one of these being the sort involved in causal relations (1986b, pp. 241ff). The convergence of a numerical sequence may, says Lewis, be an example of an event, but an event of another sort. In developing a theory of events of the causal sort, we should be guided by the causal facts. Lewis takes another step, saying he aims to develop a theory of events constrained by his own analysis of causation, a theory that allows his causal analysis to be right (p. 243). This, I take it, is at least one reason why he does not consider blaming the Dependence Thesis for the bad results above, but rather blames the alleged events. (Perhaps he also

figures that the alleged events are odd enough that we will not really miss them—all his examples of very extrinsic events are quite strange. But there are more ordinary examples, as I showed above.)

I, on the other hand, am willing to question an analysis incompatible with the view that adulteries and assassinations (construed extrinsically) are events of the causal sort. So I must ask, is the Dependence Thesis to blame? Might it not be improved? Perhaps it might. I think it holds up fine when neither of the distinct events in question satisfies some necessary criterion of the other one, as we saw that Socrates' death does for the widowing and Husband-presence does for the adultery. For those cases, the Dependence Thesis' requirement of *spatiotemporal* distinctness does not exclude them as causally connected—we would need to require something more like a "logical distinctness." But I do not know exactly how to improve on the Dependence Thesis along these lines. I shall exclude all very extrinsic events from the domain of our topic, because I (like everyone else) do not know how to include them. Including them will just have to remain unfinished business—either by special treatment or a new and improved Dependence Thesis. Or, alternatively, it will have to be argued that very extrinsic events are never causes and effects, despite appearances. Among predominantly intrinsic events, I see no trouble with the Dependence Thesis.[16]

(Lewis (p. 263) also complains about another kind of dependence situation between very extrinsic events, other than the one where one event satisfies some necessary criterion of the other's occurrence. Here is an example: If the drinking of the hemlock had not occurred, the widowing would not have occurred; so the drinking comes out a

[16] I fend off some unrelated prima facie counterexamples to the Dependence Thesis in the next chapter.

cause of the widowing. In this case, the drinking does not *itself* satisfy an essential criterion of the widowing—rather, it is a cause of one, a cause of the death. Lewis thinks that also this sort of causal result is wrong.[17] To my mind, it is not clearly wrong. Socrates' drinking hemlock was a cause of the widowing, as deciding not to sign the divorce papers was a cause of the adultery. These sound OK, to me. (Sure, there may be an implication of causation faster than light, if Xanthippe's region is very far from the drinking or Husband's decision is very far from the bedroom; but obviously there is no suggestion from science that faster-than-light causation *between very extrinsic events* is unlawful. So that is a red herring.))

A *second* restriction of Lewis's is that event essences should not be extremely "rich" (1986b, p. 256). Lewis gives as an example of an overly rich "event" the unit class of a region of the actual world. With his "part" definition, Lewis strove for the result that two events can occur in the same region yet be distinct (see footnote 10). His complaint about the unit event is that no distinct event can occur in its region. On my definition, this sort of co-occupation is not possible anyway. (And as I explain in footnote 10, on Lewis's definition it is arguably less possible than it is supposed to be, if it is even possible at all.)

But there is another problem with overly rich events, which was pointed up by Jonathan Bennett (1988, p. 65). An "event" $f$ that is a unit class of an actual region will be counterfactually dependent on every occurrent event $e$: "If $e$ had not occurred,..." takes us to non-actual worlds, and at none of them will there be a member of $f$. (Bennett

---

[17] That is not his very example, but it is relevantly akin. Lewis uses this: "If the widowing had not occurred, the cooling of Socrates' body would not have occurred." But my counterfactual is more clearly true. One way for the widowing not to occur is for Xanthippe not to be present in R—we cannot insist that

says a unit event would be dependent on *almost* every actual event—but I do not know why he says "almost," and he does not explain.) Thus, thanks to the Dependence Thesis, every event that is distinct from *f* will come out a cause of it.

This draws our attention to what is so peculiar about *f*: not only is it (metaphysically) impossible for it to occur in any other time, place or manner, it is impossible for it to occur in any other *setting*, with things being any different outside its region. Thus whether it occurs is very much a matter of what goes on outside it, in that occurrences or non-occurrences outside its region are sufficient to imply that it does not occur. *f* occurs iff the entire history of the world is as it actually is! As far as I can see, overly rich events are overly extrinsic events. But then this second restriction, against richness, can be dropped as redundant. (Excessive extrinsicness will not in turn always be excessive richness, so we can drop the richness constraint but not the other. Consider the event that occurs, in spatial region S at time *t*, iff somebody somewhere is jogging at *t*. It is far from rich, but it is very extrinsic.)

The *third* of Lewis's restrictions is that disjunctive events must not have highly varied disjuncts (1986b, pp. 266ff). Lewis defines a disjunction of events $f_1, f_2, \ldots$ as an event that occurs in a region iff one of the *f*'s occurs there. He accepts that there are some such events, such as, he says, a stamping that is a disjunction of an event that is essentially a left-foot-stamping and an event that is essentially a right-foot-stamping. I would add that we take events to be disjunctive quite regularly, since we think they could occur in a variety of ways: the party could have been upstairs or downstairs, so it is a disjunction of

---

it is the extrinsic condition (Socrates' death) that would have failed rather than the intrinsic. Thus, the cooling *might* not have occurred.

a party that is essentially upstairs and a party that is essentially downstairs. However, events, Lewis says, do not have disjuncts that are highly miscellaneous—disjuncts, he says, such as an event that is essentially a walking and an event that is essentially a talking. His reason for rejecting such events is that they will often stand in counterfactual dependence with other, distinct events when in truth there is no causation between them. Suppose Fred talks, and thereby, in the same region as the talk, there occurs a talking-or-walking. Fred's talk is a cause of Ted's laugh: if the talk had not occurred, the laugh would not have occurred. Likewise, then, the talking-or-walking, if it is a real event, causes Ted's laugh, since it will also be true that if the talking-or-walking had not occurred, (the talk and) the laugh would not have occurred. Lewis simply says this causal result is counterintuitive.

There are two reasons it might sound false to say that the talking-or-walking caused Ted's laugh, even though it is true. The first is that to speak in such disjunctions is in general odd and inappropriate. Suppose a sound is caused by a stamping that is actually with the left foot but could occur with either foot (the stamping Lewis accepts, above). Consider: "The left-foot-stamping-or-right-foot-stamping caused a sound." That does not inspire confident assent either. Suppose we give the talking-or-walking entity a nondisjunctive name. Call it a *yalking*, or a *yalk*. Some yalks are walks, some are talks, just as some stampings are left-foot stampings, some are right-foot. Now I say, "The yalk caused Ted to laugh." Is that so clearly false? I hope you agree it is at least less so. But let me turn to the other, related reason we may feel inclined to deny it.

The second reason we may want to deny that the talking-or-walking—the yalking—caused Ted to laugh is that the event is strange to us, for we have no particular interest in

grouping such disjuncts. No surprise, then, that claims involving it sound dubious. Even *non*causal ones sound wrong: "the talking-or-walking was loud" is no better than "the talking-or-walking caused a laugh." But we understand what a talking-or-walking is—it has been clearly defined for us. If we make an effort to see it as an event, which, according to the Fred story, occurred in the region of Fred, or Fred's upper body, perhaps it will seem *correct* that it resulted in Ted's laughter—just as the stamping, which occurs in just the region of the left leg, causes a sound. We can try to see it as an event by imagining a context in which people genuinely care about it and then assessing whether, in that context, it seems wrong to think of it as a cause. If it does not, we have to ask ourselves whether we want to reject alleged events on the basis of a lack of interest. Consider the following.

There is a land where talking and walking are difficult and rare, and they are considered the ugly bodily actions. Such acts are referred to, with some disdain, as "yalks." "Yalking" before strangers is considered very impolite. Yalking during church sermons is *quite* taboo. "Absolutely no more yalking in church, Junior!" she scolded. Junior eventually gets suspended from church services. His last yalk caused quite a stir among parishioners. He yalked by talking, and most folks consider talk-yalking in church to be more rude than walk-yalking (since walk-yalks are quieter). "If his yalking had been walking instead of talking, I might not have lost my temper," said the minister.

*They* certainly think yalks have effects. It seems that once one stops to take an interest in a certain disjunctive, alleged event with highly varied disjuncts, one can see it as a cause, and as a genuine event. Do we want to be rejecting alleged events on grounds of

disinterest? That is, do we want an ontology of events to be *relative* to interests (or to something keyed to interests, such as a context or a society)?

I think not, because there is a simpler option, which is just to allow miscellaneously disjunctive events and to recognize that some events we care about and some we do not. What good is the added complexity of keeping track of shifting event ontologies? If we change and come to care about yalks, we will have to say that *now*, or *for us*, or *given our interests*, it is true that Fred's yalk occurred, though back in 1986, or given our old interests, it was false. Why bother with this odd and tricky relativism?[18]

Some *sums* we care about: we think that the sum of a murder and a nearby suicide by the murderer is an event, which we call a "murder-suicide." The mereological sum of the passing of a comet and Larry's cooking bacon is something we do *not* ordinarily take for an event...well, *we* don't, but Larry does: Larry believes that cooking bacon while a comet was near caused his sterility. "The combination of the bacon-cooking and the passing comet caused my sterility," he says. "It is a kind of ice-fire-carnivore event....those are *bad*; the ice-fire-carnivores poison your karma." Larry might not be mad, he might just be badly educated; or this may be a work of fiction, in which the composite in question really *does* cause sterility! There is no particular problem with accepting unrestricted composition; it just leads to events we (or must of us) do not care about. Rather than go relativistic, it is simpler just to recognize that some sums we care about, some we do not. This defends my adoption of the mereological axiom of unrestricted composition.

---

[18] Notice we are *not* talking about relativity of an *interpretation* here; the *meaning* of a given yalk statement is the same before the interest-shift as after. We are talking about rejecting the property of being an event in favor of a relation: x is an event relative to y.

I think that, for the present project, we can simply take an *event* to be any formally eligible class of regions (as defined above) that is not overly extrinsic. (I do not have a definition of "overly extrinsic.") This will give us events that are a hodge-podge of parts, that are highly disjunctive, that have any random parts or features essentially, that could occur in any manner at all, at any time, here, there or anywhere. For example, we have the event that occurs iff some person and at least three mice or trees cough loudly or burn in Asia or on Mars during the 14th century or any time after 2001, and it essentially occurs in the union of those regions that contain the described activity. We also have the event of something happening in Europe, which occurs throughout Europe as long as something is happening there. And we have the event of something happening somewhere, which occurs throughout the world as long as something is happening (it is the class of all possible whole-world regions). To the intuitive oddness of these events we can take the stance taken above: Yes, the events are odd, but this is primarily because we are not interested in them; we can accept that they are events but of sorts we do not care about. I cannot find that such liberality creates any *problems*.

In sum, it appears that, for the present project, we need the proscription against excessive extrinsicness in constraining the formally eligible classes of regions, but no other restriction is needed.

I will not be burdening you with bizarre events in the chapters that follow (though I do hope you will not be *very* squeamish about what counts as an event). My purpose here is to be specific about what I take an event to be. This will help in considering potential counterexamples and responses to the various analyses, including my own. As we will

see, when examples get even a little tricky, intuitions can get murky, yet it is often clarifying simply to be specific about what we are taking the events in the examples to be.

# CHAPTER 3

# DEFINITE PROBLEMS FOR THE COUNTERFACTUAL APPROACH

## 3.1 Introduction

The purpose of this chapter is to home in on the definite problems for the counterfactual approach to event causation. I begin by formulating two simple analyses representative of the approach—the first is supposed to handle deterministic worlds, and the second is supposed to handle all worlds. I then defend them against some alleged counterexamples. I describe two types of genuine counterexample, representing two problems that adequate analyses must solve: the preemption problem and the problem of failed potential causes. Last, I describe alleged instances of the preemption problem that I shall not accept. Chapter 4 will examine contemporary analyses with an eye to how well they cope with the two problems.

## 3.2 Two Simple Analyses

The search for a counterfactual causal analysis begins here: Event *c* is a cause of event *e* iff, were *c* not to occur, *e* would not occur. I assume (Section 2.4) that there are nonoccurrent events (i.e., events that could occur but do not), in which case the analysis delivers nonoccurrent causes and effects: it is true that if I were not to fly to the moon now, I would not fly to the sun, but my flying to the moon is not a cause of my flying to the sun, since in fact I do neither. So let us add to the right side that *c* and *e* occur. It would also be an improvement to require that *c* and *e* be non-identical, lest we get that every event is a cause of itself ("if the explosion hadn't occurred, it wouldn't have occurred"). Further, even if *c* and *e* are allowed to be partially overlapping, we get

unintended results: that writing "y" is a cause of writing "Larry."[1] So, an improved start is this: Event *c* is a cause of event *e* iff *c* and *e* occur, *c* and *e* are distinct (i.e., non-overlapping),[2] and were *c* not to occur, *e* would not occur.

In Chapter 1 we saw this definition from David Lewis: Event *b* is *counterfactually dependent* on event *a* iff (i) were *a* to occur, *b* would occur, and (ii) were *a* not to occur, *b* would not occur. I assumed (with Lewis) that if *a* and *b* occur, (i) is true and the dependence reduces to (ii). Thus the improved start, which will be our *deterministic simple analysis (DSA)*, says that *c* is a cause of *e* iff they are occurrent and distinct and *e* is counterfactually dependent on *c* (this is "the simple analysis" of Chapter 1).

A different start is needed in order to handle indeterministic causation. If I place a radioactive sample before a Geiger counter, the chance of a click within the next minute is less than 1. If a click occurs, a cause of it is the placement of the sample. DSA succeeds in getting that result, since if the placement had not occurred, the click would not have occurred. However, consider the following sorts of case. An atom moves to a higher energy state and its chance of emitting a particle within the next minute thereby increases from .1 to .9; the emission occurs. Someone smokes a pack a day for five years, increasing his chance of getting lung cancer, and he gets lung cancer. The state change and the smoking seem to be, in some good sense, causes, yet their respective effects might still have occurred if they had not.

For this reason, the following starting point has looked better to many: Event *c* is a cause of event *e* iff, were *c* not to occur, the probability of *e* would be lower (than it

---

[1] This is a slight alteration of an example from Jaegwon Kim (1973a, p. 206).

[2] See mereological definitions, Section 2.4.2.

actually is). Such an approach is usually meant to accommodate not only indeterministic causation but also deterministic; for deterministic, the idea behind the analysis is that without *c*, *e*'s probability would be 0 instead of 1.

This second, probabilistic start is improved by the requirement that *e* occurs: even if the cancer had failed to take place, still it would have been less probable had the smoking not occurred; we do not thereby want to count the smoking a cause of the cancer that did not happen. Once this requirement has been added, there is no need to require that *c* occur—if the requirement that *e* occur is met, then if *c* does not actually occur, the counterfactual antecedent will be true and the consequent false, in which case the counterfactual will be false (Section 2.2). But there is no *harm* in requiring that *c* occur, so let us do it for aesthetic reasons.

We can make the start more precise with another notion of dependence, adapted from Lewis (1986a, pp. 176–7).[3] The notion of probability at work in the definition is that of objective, single-case probability, or "chance," which I discussed in Section 2.3. Event *b* is *probabilistically dependent* on event *a* iff, for some range *R* of chances, (i) were *a* to occur, the chance of *b* at (the end of) *a*'s time would be within *R*, and (ii) were *a* not to occur, the chance of *b* at that same time would be below *R*. (I will comment below on this choice of time for chance comparison.) I use a chance *range* mainly because, as Lewis says, sometimes there is no unique chance that *b* would have were *a* to occur.[4] I will assume that if, for some *R*, *a* actually occurs and the chance of *b* at the end of *a* is

[3] Lewis never precisely defines a probabilistic dependence relation, but readers regularly infer something along the following lines from what he says.

[4] Paul Noordhof (1999, p. 97) reminds us to make the definition sensitive to this point (though he does not use a range).

within $R$, then for that $R$ clause (i) is true. So far, then, the probabilistic start is this: $c$ is a cause of $e$ iff they occur and $e$ is probabilistically dependent on $c$.

Similar definitions have been used by others. Some (including Lewis) prefer "very much lower," though I cannot find a need for this (Lewis does not offer one). Some (including Lewis) focus on the time just after $a$; Paul Noordhof (1999, pp. 100, 104) uses the time just before $b$.[5] I do not like using the time just after $a$. One reason is that it makes the analysis just given unfit for temporally direct indeterministic causation, a kind of causation I see no good reason to count as impossible. The problem is that, in such a case, the time just after the cause is the effect's time.[6] Suppose that if $a$ were not to occur, at no time would $b$'s chance be 0 ($b$ is not counterfactually dependent on $a$); then were $a$ not to occur, $b$ still might, which means that were $a$ not to occur, $b$'s chance at the time just after $a$ might be 1 (the chance of an occurrent event, at its own time, is 1). So $a$ fails to come out a cause, even though it is one.[7] One problem with using the time just before $b$ is that an event may cause another by way of causing a temporally late essential part of it; here the cause may occur simultaneously with earlier parts of the effect, in

---

[5] Lewis does not give reasons for his choice of time. Noordhof's choice, as he illustrates, helps fend off some examples in which simple probabilistic analyses overcount causes (not enough examples to convince me it is a crucial time choice). We will return to Noordhof's approach in the next chapter.

[6] Time may be discrete, or, if not, one of the events occupies a temporal interval open on one end. For the case where $b$ is temporally extended, "the effect's time" should read "a part of the effect's time," and in order to make my case we will need to assume that every stage of the effect deterministically causes all the later stages.

[7] I also find "just after" philosophically a bit odd. If the chance, at $t$, of a future $b$ is implied by (the laws and) world history up to and including $t$, as is generally believed (see Section 2.3)—and if the part of history that is at $t$ is not a vacuous add-on such that the portion of history preceding $t$ will alone imply $b$'s chance at $t$—then events that occur *at* $t$ matter to $b$'s chance *at* $t$, such that without them, $b$'s chance *at* $t$ might be different. If $a$ occurs at $t$, then $a$ may help determine $b$'s chance at $t$, whatever that chance is; so there is no *need* to look after $t$ to capture the idea of $a$'s "making $b$ more probable." (It seems that, by looking just after $a$, it is as if one is trying to find the direct *effect* $a$ has on $b$'s chances. And that is not really our goal.)

which case it cannot contribute to chances at times *before* the effect. But the time of, or at the end of, the cause looks like an appropriate time for evaluating chance contribution in such a case.

Now we can consider whether we need a distinctness requirement. Let *e* be a writing of "Larry" that has its precise actual time essentially. At the time of the writing of "y," the chance of *e* was 1, since all of it was present and past; if the writing of "y" had not occurred, then the chance of *e* at that same time would have been 0, since at that time *e*'s nonoccurrence was assured; so a distinctness clause is needed to avoid counting the writing of "y" as a cause of the writing of "Larry." The improved probabilistic start, which will be our *probabilistic simple analysis (PSA)*, says *c* is a cause of *e* iff they occur, they are distinct, and *e* is probabilistically dependent on *c*.

I am not sure whether the two simple analyses should require that *c* and *e* be *essentially* distinct. The question has not been broached in the literature, yet it seems to me that writers almost always have essential distinctness in mind when they use the notion of distinctness. I have not found anything of clear importance that hangs on the question, but at least to fix ideas, and to take what seems the more cautious approach, let us interpret the analyses as requiring essential distinctness.

## 3.3 Some Merely Alleged Problems

### 3.3.1 Extrinsic Events

Jaegwon Kim (1973a) offers the following example to show that the analysans of DSA is not sufficient for causation:[8] "If my sister had not given birth at *t*, I would not

[8] Kim does this in the context of attacking Lewis 1973b, an analysis that takes counterfactual dependence between distinct events to be sufficient for causation (Section 4.3). Kim also offers as counterexamples

have become an uncle at *t*." What are the spatiotemporal regions of these two events? If they are the same—say, in the region of the sister at *t*—then the example does not fit the analysans, which requires distinctness. Perhaps it would be more natural to take the birth to be in the region of the sister at *t* but my becoming an uncle to be in the union of that region and the region in which I exist at *t*—but then once again the events are not distinct, but overlap. If the events' regions are distinct, then perhaps the birth is in the region of the sister at *t* and the uncle-becoming is in the region of me at *t*. The problem here is that, in order for me to become an uncle, it is *necessary* that certain events occur outside the region of my body; this makes becoming an uncle an overly extrinsic event, an event that is not just a matter of what goes on in its own region. As explained in Section 2.4, events shall be assumed to be predominantly *intrinsic* properties of their regions.

### 3.3.2 Asymmetry

As mentioned (Section 2.2), the exclusion of backtracking counterfactuals ("backtrackers") provides defense against a problem of effects being counted as causes of their causes. Let us look at that challenge and the defense more carefully.

Suppose a rock-throw is a cause of a window-break in a thoroughly deterministic setting. One might reason this way: (1) If the window-break had not occurred, it would surely have been that the rock-throw did not occur; (2) therefore, DSA counts the window-break a cause of the rock-throw. The argument is invalid because (1) is a backtracking counterfactual (if it is interpreted as a *non*-backtracking counterfactual, it is

---

cases in which the events at issue are clearly not distinct (I can only assume he took Lewis's use of "distinct" to mean non-identical instead of non-overlapping). I shall ignore those two examples (#s 2 and 3 in his paper). Another of Kim's example's (#1) involves things he admits may not be events (e.g., yesterday's being Monday), and I think that indeed they are not events, so I ignore that example also.

false): since the counterfactuals in DSA are non-backtracking, (2) does not follow from (1). With respect to PSA, the challenge posed by this deterministic example would go like this: (1′) If the window-break had not occurred, it would have to have been that the rock-throw did not occur; (2′) therefore, if the window-break had not occurred, the chance of the rock-throw at the time at the end of the actual window-break would have been 0 instead of 1; (3′) therefore, PSA counts the window-break a cause of the rock-throw. (2′) is not a backtracker; but (1′) is. Then why does (2′) follow from (1′)? It seems it does not unless it invokes a similarity relation from that same, prohibited backtracking class. But then since PSA is restricted to the other class, (3′) will not follow from (2′).

For indeterministic cases, I believe there is no asymmetry threat to begin with: chancy events can fail on their own, so there can be no justification for supposing that were they not to occur, it would have to be that their causes *also* failed, or that their causes' chances were lower. Let the rock-throw ($c$) be an indeterministic cause of the window-break ($e$), so that the chance of $e$ just before $e$ occurs is .99 thanks to $c$. Note that at the time $t_e$ at (the end of) $e$, the chance of $c$ is 1, since $c$ is past. Even on a backtracking interpretation, there is no ground for saying that if $e$ had not occurred, it would (have to) have been that $c$'s chance at $t_e$ was lower (i.e., 0, since at $t_e$ the rock-throw is in the past and its chance can only be 0 or 1)—for, since $e$ was never determined to occur, it might have failed to occur even given that $c$ occurred, which is to say, even given that $c$'s chance was still 1 at $t_e$. (Parallel reasoning holds where a different chance time is used for comparison in PSA. The chance of $c$, at *any* time, need not differ in order for $e$ to fail to occur.) For any occurrent event $e$ whose chance is less than 1 at all times before it happens, no claim

about how the past would have been different had it not occurred can be true, even in a backtracking context.[9]

### 3.3.3 Common Cause

The proscription against backtracking also solves a problem arising from events having a common cause. Assume determinism again. Suppose $c$ is a direct cause of $e_1$ and a direct cause of $e_2$, and $e_1$ is neither a cause nor an effect of $e_2$. One might reason like this: (1) If $e_1$ had not occurred, it would have to have been that $c$ did not occur, whereupon $e_2$, too, would not have occurred; (2) therefore DSA counts $e_1$ a cause of $e_2$. Since (1) depends on backtracking, (2) does not follow. Similarly: (1′) If $e_1$ had not occurred, then it would have to have been that $c$, and then $e_2$, would not have occurred; so if $e_1$ had not occurred, the chance of $e_2$ at the time at the end of $e_1$ would have been 0 instead of 1; (2′) so $e_1$ comes out a cause of $e_2$ on PSA. Again, the argument is faulty by relying on backtracking.

As before, there is no threat of a challenge of this sort in a chancy setting. Even under a backtracking interpretation, the needed counterfactuals are false: for instance, it is not true that if $e_1$ had not occurred, it would have to have been that $c$'s, or $e_2$'s, chance was lower (at $e_1$'s time)—for $e_1$ can fail with the chances of $c$ and $e_2$ being just what they are at all times.

---

[9] Perhaps in practice we accept backtrackers, in backtracking contexts, even while believing that the effects at issue are chancy: one might say, "If the doorbell had not rung, it would have to have been that I did not press the button." I think that if the ringing is genuinely chancy, this counterfactual is false. If the doorbell had not rung, it might have been that still the button was pressed, yet the bell failed to ring—if you do not believe that, then you are not *really* thinking of the ring as chancy. Or, you believe that as long as the button is pressed, the bell *would* ring, given *how very improbable* the bell's not ringing would be. On the

### 3.3.4 Standard Non-preemptive Overdetermination

In instances of deterministic overdetermination of an occurrent event *e*, multiple distinct events occur such that, were none of them to occur, *e* would not occur, yet *e* would still occur as long as any of the multiple did. In some such cases, some of the multiple are causes and some are not—these are cases of preemption. I will address preemption later in this chapter (Section 3.4.2). In other instances of overdetermination, either all of the multiple events are causes or none of them are—though it may be unclear which of these is the case. This is *non*-preemptive overdetermination.

Cases also divide according to whether they are *intuitively* asymmetric—that is, according to whether they inspire any inclination to count some of the multiple as causes but not others. Here is an example of the intuitively *symmetric* sort of case: two (or three, or ten) assassins' shots hit the mark squarely, such that either shot by itself would have been enough to cause the death. Clearly, each shot is a cause of the death or neither is, since neither shot is in any way special relative to the other. All intuitively symmetric cases are uncontroversially non-preemptive. Intuitively *asymmetric* cases can be controversial; many are generally accepted to be preemption cases, but some cause disagreement. In Section 3.4.2.2, we will look at controversial cases. Here I want to address the intuitively symmetric, clearly non-preemptive cases.

There is indeterministic overdetermination as well. Here, the chance of *e* is overdetermined: the chance of *e* would have been lower had *none* of the multiple occurred, but it would have been the same as long as any of them occurred. Due to the

---

incorrectness of saying that something improbable would not happen just because it is so very improbable, see the end of Section 2.2.

complexities of indeterministic causation, of which we will see more later on, the causal relationships between these events and *e* could be of various sorts. All I want to address here are the intuitively symmetric, non-preemptive cases. Example: two switches are simultaneously flipped and each flip is enough to make the chance of the lamp coming on what it was (at the time of the flips)—.99, say—and the lamp does come on.

I shall refer to intuitively symmetric, non-preemptive overdetermination, whether deterministic or chancy, as *standard non-preemptive overdetermination*.

For the assassin example (deterministic), DSA says that neither overdeterminer is a cause: for each shot, the death would have occurred without it. PSA says this also, since for each shot, the chance of the death would have been the same (=1), at the relevant time, had it not occurred. In the lamp example (chancy), PSA gets that neither flip is a cause of the lamp coming on, since for each, without it the chance of the lamp coming on would have been the same. Are these the right results?

There is general agreement that intuitions waiver as to whether this sort of overdeterminer is a cause. We are inclined to feel that in a sense it is and in a sense it is not. Some writers lean towards the affirmative; I am undecided. I shall heretofore be assuming that, as far as we know, it is acceptable not to count standard non-preemptive overdeterminers as causes at all. I shall not, in my own analyses (Chapters 5 and 6), be offering any way to count them.

Does excluding them leave a gap in the causal history of the overdetermined event? No, because we may say that some composite of the overdeterminers caused the effect: if the composite had not occurred, none of its parts would have, and so the effect would not have occurred. This claim that none of the composite's parts would have occurred if the

composite had not has been attacked, by Michael McDermott (1995a, p. 135) and Stephen Yablo (2002, p. 139): *might* not *some* of the parts still have occurred? The answer depends on what is meant by "the composite." Suppose the composite can occur only if all its parts do—each part is essential. For example, we might take "the 21-gun salute" to have all its parts essentially. I agree that it sounds fine to say that if the 21-gun salute had not occurred, *some* of the guns might have fired. And it is consistent with the occurrence conditions of the whole event, since the nonoccurrence of a single part suffices for the nonoccurrence of the whole. But suppose the composite can occur as long as *any* of its parts do—each part is sufficient. Then in supposing away the whole, we must suppose away all the parts. For example, we might take a ceremonial shotgun blast to be an event that occurs as long as at least one barrel fires, though in fact both fire. "Had the ceremonial shotgun blast not occurred, one of the barrels might still have fired" is false. As long as "the composite" means a sum for whose occurrence any of the parts is sufficient, the effect will depend counterfactually on the composite. The aforementioned attackers simply do not take this sort of sum seriously. McDermott does not consider it. Yablo simply asserts that we do not accept such a sum as an event (and he erroneously cites Lewis for support—in fact Lewis never says that there are no such sums (see Section 2.4.3 on disjunctive events, or Lewis 1986b, p. 266, last full paragraph)).

### 3.3.5 Meta-prevention

Sometimes an event $e$ depends on another $c$ because $c$ prevents some third event $d$ that would have prevented $e$—$c$ prevents the prevention of $e$. Thus I murder the clumsy trapeze artist by untying his safety net as he falls: the death depends upon the untying,

thanks to the fact that the untying prevents a later presence of net from preventing the death (again, like many, I count object presences as events). DSA counts the untying as among the causes of the death (mind you, it does not say it is *the* cause). Likewise for PSA: whether the situation is chancy or deterministic, if the untying had not occurred, the chance of the death would have been less probable, at the time at the end of the untying. This seems like a fine result, to me, but as we will see, some would disagree with it. First, consider more evidence that meta-prevention counts as causation.[10]

The removal of a pin from a grenade will cause an explosion; the destruction of a dam will cause a flood; and the cutting of a climber's rope, the opening of a gas valve, the placing of a pillow over someone's face, and the pressing of a button that merely opens a trap door and allows a bomb to fall out of a plane all have serious effects. These causes succeed by way of preventing that which could have prevented the final effect—the pillow denies one a breath, the cutting prevents a complete rope, the destruction of the dam removes a dam. This structure, it seems to me, is not hidden from view in these cases, but rather is something intuitively apparent; and without reservation we speak as if it is causal. The structure is also common and commonly perceived to be causal: we regularly think that removing, releasing, unveiling, unblocking, untying, disconnecting, letting go, cutting loose and turning off have effects—releasing the rope will cause her to fall, untying that knot will cause the tent to collapse, unveiling the statue now could cause quite a stir, turning off the power will cause the emergency lights to come on. So

---

[10] Meta-prevention is variously referred to in the literature as "double prevention," "disconnection," and "the prevention of prevention." Of these three, I think the last is most apt, but it makes a clunky adjective.

Awareness of the structure of meta-prevention, as well as an appreciation of the plausibility of the idea that meta-preventers are causes, appears in McDermott 1995b, p. 529. But it is strongly suggested in David Fair 1979 (Sections VIII and IX).

DSA's verdict on meta-prevention cases is strongly supported by ordinary (thought and) usage.[11]

And it is an important verdict, because meta-preventers are often the only causes we are interested in, and it would be hard to do without them. When the chandelier falls, we do not want to hear that the fall was caused by a gravitational field (we already know about that)—we want to hear that the chain was cut. When we want to know the causes of the lake water flooding the town, we do not want to hear about gravity or the height of the lake—we want to hear about the explosion at the dam, or the opening of too many flow valves. Likewise for many examples. (Often we are not expecting the effect *e*, because the status quo would not lead to it; then when *e* occurs, we are interested in the violations of the status quo, and in the causes of those violations—the missing net, the missing rope, pin, dam, breath of air, trap door, and the causes of those absences—and often the causes are meta-preventers.)

Thus, if we reject meta-preventers as causes, it appears we reject much confident and important ordinary talk. I know of no confident and important *denials* of causation in meta-prevention cases such that in accepting meta-prevention as causation we must still reject confident and important ordinary talk. There are just the cases, which I will illustrate below, where we are a little reticent to claim causation, but we can still muster sympathy. So it appears there is a real imbalance here, such that accepting meta-prevention as causation is preferable to rejecting it.

---

[11] See Jonathan Schaffer (2000c) for more good examples—though I think their meta-preventive structure is not ordinarily obvious without examination, which is why I have pressed other examples here. I want to emphasize that meta-preventive causation is not alien to our *concept* of causation; we claim causation even when we see the meta-preventive structure right off.

Ned Hall (2000) offers the following as a meta-prevention example for which we are uncomfortable claiming causation (and rightly so, in his view). Suzy is on her way to throw a water balloon at a dog. Billy is running at Suzy to intercept her before she does it—and he would have succeeded if he hadn't tripped. The wet dog yelps. If Billy hadn't tripped on the root, the yelp wouldn't have occurred, but it seems a little odd to say that Billy's trip is a cause of the yelp.

First, we should tweak the example. While it is true that, as Hall says, if Billy had not tripped, the yelp would not have occurred, it is not clearly true that if Billy's *trip* had not occurred, the yelp would not have occurred. Only the latter counterfactual is at issue. What is the trip? Picture it: a running boy kicks a root and goes forward and down. I think that if *this* had not taken place, Billy *might* have been running toward Suzy, but also he might have been stopped, or walking. To suppose away the trip is to suppose away a stage of a journey of a moving object, and I don't think we can say that without this stage the object *would* have been moving in a certain different place and manner.[12] The example is clearer if we put the disrupter outside the process that is being disrupted: Amy leaps at Billy and tackles him. If the leap hadn't occurred, the yelp wouldn't have occurred. DSA counts Amy's leap as a cause of the dog's yelp. That, to my ear, sounds less odd than counting the trip a cause; anyway, it may still, I take it, be something that we are reluctant to accept, something that seems more off than, say, the claim that the explosion of the dam was a cause of the flood.

---

[12] Similar remarks apply to Hall's other meta-prevention example in that paper. The candidate cause is a group of men "stopping into a pub." Hall says, if they hadn't stopped in, they would have kept going. Fine, but irrelevant. The question is what would have happened if *the stop* hadn't occurred. What is the stop? As I picture it, some men walk into a pub (what else?). It isn't clearly true that if *that* hadn't gone on, those men would have been on their originally intended course.

But as I said, this reluctance does not carry much weight once we take into account that with a little effort we can muster a good amount of sympathy for the claim after all. Imagine you expect Billy to stop Suzy from soaking your dog—perhaps Suzy attempts every day to soak it, and Billy always stops her. Today, to your surprise, Suzy succeeds, and your poor dog yelps. You learn that Amy wanted the dog to yelp, so she leapt on Billy to that end. Naturally you hold Amy partly to blame for the dog's misery and its yelping in fright. Notice that Amy's leap helps explain to you why the dog yelped; notice that Amy's leap was an effective means to the yelp; notice that it would be a bit of a strain to blame Amy for the dog's frightful yelp while maintaining that her action was in no way a cause of it. In light of this, counting the leap as among the causes of the yelp seems quite acceptable.[13]

On balance, when we survey a variety of examples (which Hall, by the way, does not do), counting meta-prevention as causation seems the better way to go.[14]

Hall complains that counting meta-prevention as causation is inconsistent with various intuitions that we have about causation—most notably (for Hall) the intuition that causation is transitive,[15] but also the intuition that, in the actual world, no event causes another at a distance without causing events in between, and the intuition that "the causal structure of a process" is determined by its intrinsic, noncausal character plus the laws that govern it (p. 217). Hall thinks that, in conjunction with our reluctance to posit

---

[13] I am not claiming that there are decisive correlations between causes and means, explanations, and objects of blame. I am just pushing intuitions here.

[14] Hall accepts that there may be a kind of "causation" on which meta-prevention counts as causation, but he thinks this must be an *unordinary* and "second-class" kind of causation (2000, p. 202). Obviously this will not satisfy the above evidence, which, as I have stressed, involves ordinary and common talk.

[15] I won't go into his examples here. We will see examples like them in Section 5.4.2.

causation in some meta-prevention cases, these violations indicate that meta-prevention is not really causation. But even if he is right about all those inconsistencies (though I am not saying he is), there is still good reason not to give this any decisive weight. For it is reasonable not to let theoretical intuitions about the nature of a subject overturn good evidence from ordinary usage, given that one's project is the analysis of an ordinary concept.

The wedding's not occurring caused much weeping; his failure to brake caused a crash; the absence of the predicted rain was a cause of the brushfire; the clock alarms failing to sound caused the participants to be absent: there are many cases in which it seems correct to talk in terms of causation by absence, or omission, or nonoccurrence. Ultimately sense must be made of this talk. I will not attempt to do it in this dissertation. However, throughout, I shall talk of nonoccurrences as causes and effects where it seems natural and correct to do so. Cases of meta-prevention raise the first opportunity to talk about causation by nonoccurrence. The cases reveal that *sometimes one event is a cause of another in virtue of a causal intermediary that is a nonoccurrence*: placing the pillow on the face causes a nonoccurrence of breathing which in turn causes the death. Intuitively, this is *how* the smothering causes the death, by way of this intermediate nonoccurrence. Similarly for all cases of meta-prevention. This is a simple but important fact about event causation that I will occasionally draw attention to in later chapters.

### 3.4 Genuine Problems

Now we turn to two genuine problems for the simple analyses. The first applies only to PSA, the second to both PSA and DSA.

### 3.4.1 Failed Potential Causes

Two radioactive samples are placed near each other, one meter away from a Geiger counter. Each placement of a sample is such that, were it not to occur, the chance of the Geiger counter going "click" within the next minute would be lower (at the time the placements occur) than it actually is, because there would be less radioactive material. One "click" occurs within the next minute. Each sample-placement made the click more probable, but only one was a cause of it, because the click resulted from a particle emitted from one sample but not the other. This illustrates the problem of failed potential causes, events that at some point from the time they occur stand a chance of being causes of a certain event but in the end do not succeed, though the event still occurs (by other causes); often such events make the effect more probable.[16] The problem is that PSA counts these non-causes as causes. There is no problem here for DSA, since the effect will not be counterfactually dependent on such a non-cause—without it, the effect might still occur, thanks to the other potential causes. (Clearly the problem of failed potential causes arises even in thoroughly indeterministic worlds.)

Failed potential causes may be divided into two kinds. In the kind just illustrated, the failure is the result of the failure of a causal intermediate: one sample-placement fails to cause a particle emission, which in turn fails to trip the Geiger counter. Another kind is pointed up by Jonathan Schaffer (2000a)—he calls it "overlapping"—and it need not involve a failed intermediate. Here is one of Schaffer's examples. Spells PK and PQ are cast simultaneously. They are of different types: PK generates a .5 chance of the prince

[16] This sort of problem for the probabilistic counterfactual approach was pointed up by Peter Menzies (1989a), though he (1989a, 1996) casts it as a problem that arises only in preemption situations (described below), which is not accurate. The example just given is not a preemption situation.

and the king becoming frogs, while PQ generates a .5 chance of the prince and the queen becoming frogs; the occurrence of each transformation *pair* is, in any situation, probabilistically independent of the occurrence of the other and of the occurrence of spells cast upon another pair; and neither spell generates any chance at all that *only one* of its potential victims becomes a frog—it is both or neither. So the prince's chance of becoming a frog is .75. As it turns out, the king and prince, but not the queen, turn into frogs. Then it seems only PK caused the prince to become a frog, even though each spell made it more probable (without it, the chance of the effect at the time of the spells would have been lower than it actually was).

Does this challenge apply to thoroughly indeterministic worlds? I think it does. Add that the frog transformation of each royal person always has a slight independent chance. (We can still leave it that a spell makes the chance of its corresponding pair .5.) Suppose that without either spell, this background chance is .01 for each person. Now an effect pair will be independent of the other kind of spell and of the other pair *modulo* the independent background chances: for example, with no spells, the chance of the queen and prince both becoming frogs is .0001; add PK and it must go up, since PK makes the prince transformation more probable—it goes up to (.5 x .01) + (.5 x .0001), or .00505. If both spells are cast, the chance of the prince becoming a frog is still .75. If only the prince and king change, it seems that, since PQ contributes only to the chance of the prince/queen *pair* (and what it contributes to the other pair is only a byproduct of contributing to the prince/queen pair), it did not help—did not cause—the prince to become a frog.

### 3.4.2 The Preemption Problem

#### 3.4.2.1 Description and Simple Illustration

Suppose a cause of *e* prevents, or at least has some chance of preventing, another event from being a cause of *e*. In many such cases, were the cause not to occur, *e* might (or would) occur, or its chance might be lower. Where the prevention goes through, that is preemption—a potential cause is preempted by an actual one—but even the probabilistic potential of preemption can make trouble for PSA, as we will see. Let *the preemption problem* be that in virtue of (potential) preemption, analyses fail.

Independently, you order Roger to go ring the breakfast bell and I order Dolly to do it. Dolly is very eager, and she pushes Roger aside on her way up the bell tower; consequently, she rings the bell and Roger does not. The two orders overdetermine the ringing; but only my order causes someone to go ring the bell, so it seems my order is a cause of the ringing and yours is not. The problem is that if my order were not to occur, your order would cause the ringing, since Roger would follow through. My order preempts your order, and consequently DSA fails. A "neuron diagram" of the sort Lewis uses (1986a, p. 200) will be helpful:[17]

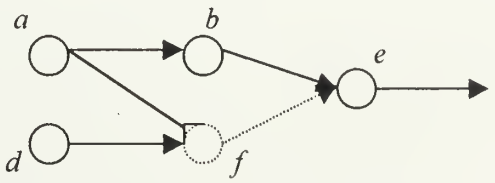


Figure 3.1

[17] Each circle-with-arrow(s) is a "neuron." Solid neurons fired, dashed ones did not. Normal arrowheads stimulate, reverse arrowheads inhibit. By law, neurons will fire if, and only if, stimulated, unless they are simultaneously inhibited. The system here begins on the left, upper and lower processes running synchronously. Letters label the firing events.

Although it seems *a* is a cause of *e*, the backup process spoils *e*'s counterfactual dependence on *a*: without any such alternate process in the picture, *e* would be dependent on *a*; but as it stands, were *a* not to occur, the backup process would go through and *e* would occur anyway. The diagram can also represent (stages of) the breakfast-bell story: *a* and *d* are my and your orders, respectively; *b* is some late stage of Dolly's journey; *e* is the ringing.[18]

Now let us suppose the situation is indeterministic. Specifically, let the backup process be very reliable, while the successful process is very unreliable, save for the inhibitory power of *a*, which is at least moderately reliable (Menzies, 1989a, p. 646). Then *a* makes *e* less likely instead of more, by increasing the probability that only an unreliable process will be in place after *a* occurs—*e*'s chance at the end of *a* would be higher were *a* not to occur. So PSA misses counting *a* as a cause. (One can see that this problem, too, arises in thoroughly indeterministic worlds.)

Notice that if the inhibited neuron had beaten the odds and fired anyway, so that all the neurons in the diagram fired, then still *a* would have failed to contribute to *e*'s chance at the end of *a*'s time—for whether the inhibited neuron fires can have no bearing on chances at times *before* it.[19] We see, then, that even when the inhibition succeeds, it is not in virtue of preempting *d* that *a* fails to be counted a cause; rather, it is in virtue of *a*'s

---

[18] (I am assuming that in supposing away *a* we suppose that *neither* axon carries a signal. I think there are common-sense identity conditions for *a* that demand this—but I will not defend that here. One always has the option simply of tacking on a stimulatory neuron before *a* and letting *that* be the preempting event.)

[19] I am assuming that if the inhibited neuron fired and *b* occurred, each of those events would contribute to *e*'s chance, as potential joint causes of *e*; in that case, at least, it is clear that *a* is a cause of *e* when both processes go through. (Where instead there is overdetermination of *e*'s chance by the two direct firings on it, it is not clear to me that *a*, or any other single firing, should count as a cause of *e*.)

significant *potential* to preempt *d*—a relatively potent potential cause of *e*—that *a* fails to be counted a cause.

We should consider the case where chances are extreme, all 0s and 1s. In such a case, clearly *e*'s chance at the end of *a* would still be 1 were *a* not to occur.

Suppose the depicted situation is chancy throughout. Then were *d* not to occur, the chance of *e* at the end of *d* would be lower (Menzies, 1989a, pp. 646–7). So *d* is wrongly counted a cause by PSA. It is important to see that this spurious causation is not part of the preemption problem—*d* is counted a cause *not in virtue of* (potential) preemption. Rather, this is just an instance of the problem discussed in the previous section, that of failed potential causes. Notice that *f* need not be inhibited by a *cause* of *e* in order to fail to occur; nor, in fact, need it be *inhibited* at all, since the situation is chancy—*f* can simply fail on its own to occur. But in these other situations, still *d* contributes to *e*'s chance and is wrongly counted a cause of *e*. The fact that *f*'s failure was made even more probable by an inhibitor, and a preemptor at that, seems to add nothing of significance to the problem. *f* failed because it was not determined; this would be so with or without *a*. We can expect that a decent solution to the problem of failed potential causes will be general enough to solve it for the cases that do involve preemption; so there is no value in making the preemption problem's solution more difficult by adding to it failed potential causes that are also preempted.

There are other, interestingly different varieties of preemption, but we will take these up in the context of proposed solutions in the next two chapters.

In Figure 3.1, the backup causal sequence, or process, fails to go through, since *f* does not occur. As I shall push, *all* clear cases of preemption involve a backup sequence that,

thanks to the preempting cause, does not completely come off—at the least, there are possible parts of events in the backup sequence that are prevented by the preempting cause. Several alleged examples of preemption have been offered that challenge this *Failure Thesis*. I shall now discuss these and explain why I do not accept them as genuine preemption examples.

#### 3.4.2.2 Alleged Preemption Set Aside

##### 3.4.2.2.1 Allegedly Preemptive Prevention

If I stop a baseball with my hand before the ball hits a window, the presence of my hand is one of the causes of the later presence of intact glass in the window (the latter is dependent on the former). It is a meta-preventive cause (Section 3.3.5). Now suppose that Ann's raised hand was between me and the window and would have stopped the ball had I missed it; we might *still* want to say that my hand's presence was a cause of the later intact-glass presence (John Collins 2000, p. 223).[20] That is, we might just want to count the presence of Ann's hand as a backup potential cause that was preempted by the presence of my hand. But then, the problem for the Failure Thesis is that the backup process actually goes through just as it would without the (putative) preemptor: in both cases we have the presence of Ann's hand, the nonoccurrence of the ball flight between her hand and the glass, and the presence of intact window.

Picture yourself in a special suit of 100%-reliable, bullet-proof armor. Someone shoots at you (in the direction of your heart) with a .22-caliber pistol, and I, always trying to play the hero, dive in front of you and take the bullet. Did I save your life? Offhand,

---

[20] The example is a variant of the wall-backup version described below, which is due to Michael McDermott (1995b).

"No" seems like the better answer. For given your armor, you were in no danger from that bullet. But take off the armor and my leap clearly *does* save your life. Whether the armor is present makes a strong intuitive difference; and the corresponding shift in judgement as to whether the leap saves your life (is a cause of your later live presence) seems appropriate. Now why was the corresponding shift not as clearly appropriate in the previous story? That is, why was it more plausible that the first hand saved the glass both with and without the backup? We must consider some other examples.

If behind my hand was, instead of a second hand, a "high, broad, thick, sturdy wall" (Lewis 2000, p. 197), then we would, I think, be a little less inclined to count the first hand's presence a (preempting) cause of the intact-glass presence, though perhaps we still might. If I were to block a lightly tossed, live hand grenade whose velocity vector at the moment I stopped it was pointed at the future location of a flying eagle five miles off, we would surely *not* be inclined to say that my hand's presence saved the eagle's life, or was a cause of its later normal flight—here a backup is the gravitational field that would have pulled the grenade down if my hand had not been there (this example is adapted from Collins 2000, p. 224). Picture a bowling lane with a high wall of sandbags across the middle of it. If I were to stop a ball that was rolling at a modest speed down the beginning of the lane, would we say that I saved the pins from falling, that my act was among the causes of the pins standing up shortly thereafter? Not likely. In sum, as I see it, we have the grenade example and the special-armor example which we are very disinclined to call preemption; we have the bowling example where we are pretty disinclined, the sturdy wall example where we are a little disinclined but somewhat on

the fence, and the two-hand example where we lean towards calling it preemption. How can DSA reconcile this dynamic of intuitions?

I suggest that *whenever* there is this kind of backup disrupter in place, the first disrupter is not a cause of the final event, and the main reason it sometimes seems better to say otherwise is that, as has often been noted, we have a strong and pervasive tendency to view the world as indeterministic. Surely this tendency is stronger the more the situation involves human action and projectiles in the air. Notice that for the kind of case at issue, if all *is* chancy, then the first disrupter makes the final event more probable, because there is a chance the second disrupter will fail to stop the process that threatens the final event. If the first disrupter makes the final event more probable, this may make it look like a cause, and indeed it may be one. It is plausible that the stronger the temptation to view one of these situations as thoroughly chancy, the stronger the temptation to count the first disrupter a cause. But since DSA applies only to deterministic settings, we ought to set aside any causal intuitions that arise only from a sense that, contrary to stipulation, the situation is chancy. When I said, "special suit of 100%-reliable, bullet-proof armor," this made intuitively clear the extreme improbability of the projectile getting by the second disrupter, and the first disrupter did not seem like a cause. To the extent I embrace the fact that there is really zero chance that the ball will pass or penetrate the wall if the first hand is absent, the catch seems to achieve nothing. As hard as it is intuitively to embrace extreme chances where we are not used to doing it, it seems to me that, to the extent I do it, the first disrupter in all such stories is no more a savior than my leap when I take a bullet in the armor story. As for the eagle example, plausibly the catch so clearly does not seem like a cause because we are inclined to think

that there is no chance the grenade will reach the eagle—we are in the habit of taking the earth's gravitational field as a definite given, something with no chance of vanishing.[21]

Collins (2000) and Lewis (2000) (and perhaps Hitchcock (2001)) want an analysis to be flexible in a way that respects our varying intuitions about these cases. Their general idea is that whether the first disrupter is a cause is relative to the remoteness of the possibility that the threat process, from a point between the two disrupters, could proceed beyond the second disrupter and prevent the final event—if very remote, the first disrupter is not a savior (not a cause); otherwise it is. How remote is "very" remote is something we are not set on: our mood varies, and contextual factors may help to (roughly) draw the line. I have no reason to think this tack cannot succeed, and I have said nothing that tells against it. My point is that there is another answer that appears satisfactory, and it leaves SCA in the clear: it is the range of our inclinations to view the cases as chancy that explains the range of our causal intuitions; and since the cases are stipulated to be deterministic, causal intuitions resulting from viewing them otherwise need not be respected.

3.4.2.2.2 Paul's Bullet

L. A. Paul (1998a, pp. 52–3) puts forward the following as a case of preemption.[22] The final effect ($e$) at issue is the presence of a bullet in spatial region S. As it happens, $e$ occurs at $t_1$. $e$'s occurrence conditions allow that $e$ could occur later, at least as late as $t_2$,

[21] On reflection, some may be able to locate a feeling that there is *some* non-zero chance of $e$, for *any* future potential $e$. But I don't think this is an intuition normally in play, and anyway if one truly accepts that there is some chance that the gravitational field will vanish, I think one should be happy to say that the grenade-catch was a cause.

[22] The version of the example I discuss comes from a suggestion in her footnote 7.

and that it could involve any bullet. The bullet whose presence actually constitutes $e$ is in flight, so that a cause of $e$ is the same bullet's presence immediately beforehand ($c$). A second bullet is not far behind, passing through S at $t_2$; and so if $c$ had not occurred, $e$ would have occurred anyway, at $t_2$, in virtue of this second bullet's presence. (So there are actually two distinct presences of a bullet in S, both presences falling within the temporal boundaries of $e$'s occurrence conditions.) In this example, none of the backup process is missing, in apparent violation of the Failure Thesis.

But I think $e$ is overly extrinsic. As discussed in Section 2.4, I take events to be predominantly intrinsic properties of spatiotemporal regions (following Lewis). Very extrinsic events are set aside as outside the scope of the present project of causal analysis. Here is why we should count $e$ as overly extrinsic.

First notice the trouble that $e$ generates. In addition to $e$, let us talk about the two occurrent presences of a bullet in S that essentially occur at the times they actually occur, $t_1$ and $t_2$: call these events $e_1$ and $e_2$ ($e_1$ and $e$ actually occur in the same region, while $e_2$ occurs in another region). The claim is that if $c$ had not occurred, $e_1$ would not have, though both $e_2$ and $e$ would have. It seems pretty clear, then, that if $e_1$ had not occurred, $e_2$ and $e$ still would have. But then it appears that if $e$ had not occurred, $e_2$ would not have occurred—for in supposing away $e$, it is not enough to suppose away $e_1$ and leave $e_2$. Now if counterfactual dependence between *actually* distinct occurrent events is sufficient for causation, $e$ gets counted a cause of $e_2$, which obviously it is not (remember that $e$'s region is the same as $e_1$'s). But as I said above (Section 3.2), it may be that it is only dependence between *essentially* distinct occurrent events that is sufficient for causation. But even then there is still a problem: $e$ may wrongly get counted a cause of

some direct effects of $e_2$; for it may be that some such effects would not have occurred if $e$ had not occurred, since in supposing away $e$ we suppose away $e_2$. The problem arises because in order to suppose away $e$ we have to suppose away an event actually distinct from it. This is because $e$'s occurrence is implied by goings on outside its region—by $e_2$. Where an event is implied by goings on outside its region, it cannot be said to be intrinsic to that one region, but largely extrinsic to it.[23]

3.4.2.2.3 "Rumpelstiltskin II"

This is from Michael McDermott (1995b, p. 534—the *second* Rumpelstiltskin story in the paper). He does not advertise it as specifically a preemption-type challenge, but it is one. I ask the prince his name. One witch has cast a spell that will turn him into a frog if he answers "Rumpelstiltskin," and a second witch has cast a spell that will turn him into a frog if he does *not* answer "Rumpelstiltskin." The prince answers "Rumpelstiltskin" and turns into a frog.

A stated intention is that the spells and the saying work directly, so that there are never any causal intermediates between them and frog transformations.

McDermott says that the saying of "Rumpelstiltskin" is a cause of the prince's transformation. Yet had the saying not occurred, the effect still would have. It seems, and McDermott claims, that the second spell is not a cause; notice that if the saying had

[23] It may be that *e occurs iff* $e_1$ *occurs or* $e_2$ *occurs*. There is nothing wrong with that disjunctiveness, for it is true generally that events that can occur in various ways have an occurrence condition that is a disjunction of occurrences of the specific possible variants (see Section 2.4.3). The real problem is that the *regions* of occurrence do not track the disjuncts: When both $e_1$ and $e_2$ occur, $e$ is in $e_1$'s region only. The occurrence condition just given is silent on $e$'s region of occurrence—that does not make it false, but in order for $e$ to be a genuine event, it must *also* be true that $e$ occurs only if $e$ occurs in the union of the regions of whichever of $e_1$ and $e_2$ occur. Otherwise, that $e$ occurs can be implied by goings on outside $e$'s region, and $e$ is thus not intrinsic enough to its region.

not occurred, the second spell would definitely have been a cause, one on which the effect was counterfactually dependent. Therefore, the example threatens to be an instance of preemption, where the saying preempts the second spell.

I think there is an intuitive temptation to view the saying as somehow "cutting off" or "deactivating" the second spell. But it is the author's intention that there is no such failure in or between the second spell and the effect. A "neuron" example, using analogous laws, may be clearer; the clearer is the intended structure, the more my response can be appreciated, so let us consider the following.

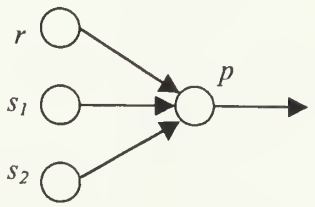


Figure 3.2

$r$ is the firing of an R-type neuron, $s_1$ of an $S_1$-type, etc. By law, a P neuron will fire if fired upon simultaneously by $S_1$ and R neurons, and, by law, a P neuron will fire if fired upon by an $S_2$ neuron while no R neuron also fires upon it. All neurons fire. If $r$ and $s_1$ cause $p$, and $s_2$ does not, then $r$ preempts $s_2$—preemption without any failure in the backup process.

McDermott's example has been ignored. No counterfactual analysis (save perhaps the one McDermott proposes in the paper) counts $r$ but not $s_2$ as a cause.[24] But the example

[24] McDermott uses a somewhat complex maneuver (based on counterfactual sufficiency rather than necessity) that I will not go into. Most analyses count both or neither of $r$ and $s_2$ as causes; my defense will imply that counting both or neither is acceptable.

has intuitive pull; some response is called for. There is a response that I hope you find good enough to justify setting the example aside as being not clearly successful. Fortunately, an analogous response will apply also to allegedly preemptive "trumping" cases, which I take up in the next section.

Let C represent a thorough characterization of the background circumstances of $r$ and $s_2$: the presence of the P neuron in a suitable environment (wet, pressurized, warm, etc.), the absence of potential inhibitors of $p$ (inhibitory neuron firings, meteorites, etc.), the absence of other neurons connected to the P neuron other than those shown in Figure 3.2, and the occurrence of $s_1$.[25] C does not lawfully imply (the occurrence of) $p$. But C plus $r$ does, and, less obviously, C plus $s_2$ does (various derivative laws make this so—for example, "If $S_1$ and $S_2$ neurons fire on a P neuron, the P neuron will fire"). Thus $p$ is actually implied by C plus $r$ but also by C plus $s_2$. This double implication is *not obvious* at first; but when I look at Figure 3.2 and attend to it, the case looks to me a lot like a case of standard non-preemptive overdetermination: in the example in which two assassins hit their target at the same time, the background circumstances plus either shot implies the victim's death, just as C plus $r$ and C plus $s_2$ both imply $p$. And when I see the case in this light, I have to some extent the same reaction that I have generally for standard non-preemptive overdetermination: for each candidate cause, it seems it's helping *and* it seems it's not—it seems it's *not*, I think, because it is occurring in surroundings that intuitively are already enough for the effect, namely the background circumstances plus the other candidate—and consequently it seems acceptable to count each candidate as a

[25] Throughout, when I say "the occurrence of $e$," I mean *that* e *occurs*.

cause but also acceptable to count neither. The upshot is that I can see the case as not preemption, and as one for which DSA gets an acceptable result.

(The ways C plus $s_2$ and C plus $r$ imply $e$ are not quite symmetrical: arguably C plus $s_2$ implies $p$ only with laws that are derivative. But the intuitive question remains: isn't C plus $s_2$ *enough* for $p$? If the answer is not clearly Yes, still it seems, to me, it is not clearly No, either. For some derivative laws, the antecedent situation implies the consequent situation by way of implying some third situation (a common cause, perhaps): e.g., falling barometers are followed by falling rain. We do not think the falling barometer is enough for the rain, and it isn't—we need that third thing, too, namely low air pressure. But no such back-door implication is involved in the case at hand: $s_1$ plus $s_2$ does not imply $p$ by way of implying some other event that causes $p$. So we should not dismiss the question of whether C plus $s_2$ appears to be enough for $p$ simply because derivative laws are involved.)

When the two assassins hit their target, we blame them for the death. We say that they are responsible, and we feel inclined (rightly or wrongly) to say that *each* is responsible, that each is a murderer. If the bullet of one assassin were to arrive first, knocking the victim down so that the second bullet missed, we would hold only one assassin responsible for the death (even if we would punish the other for his attempt); this assassin would be convicted of murder, and the other assassin of *attempted* murder. Thus our intuitions about responsibility vary depending on whether we perceive an overdetermination case to be preemptive. Suppose $p$ is the prince's undeserved death. Mr. R and Ms. $S_2$ both want that $p$ occur; they know that a process has been set in motion that will bring about $s_1$ at $t$; they know all the relevant laws (including derivative ones);

neither knows or cares what the other is planning to do. Shortly before $s_1$, each orders his/her servant to bring about $r/s_2$ at $t$, and, as each knew would happen, the prince dies. I think a jury would hang both Mr. R and Ms. $S_2$ for murder. I would. The jury would be loathe to accept that Ms. $S_2$ lucked out, that thanks to the other process she is guilty only of attempted murder. *Perhaps* the jury would be wrong here, but note this contrast: if Ms. $S_2$'s servant had held back once she saw that Mr. R's servant was going to go ahead (standard preemption), then the jury would be *quick* to accept that Ms. $S_2$ lucked out, that she was not guilty of murder, but of attempted murder. Why the contrast? The moral setting seems to bring out the overdetermination take on the case. Consequently it is not clear to me that the example counts as preemption.[26]

3.4.2.2.4 Trumping

Suppose that, by law, the prince's fate corresponds to the spell cast by the tallest person, and warlocks are taller than witches—warlocks "outrank" witches, as I'll put it. Thus, if a witch casts a goat spell while a warlock casts a frog spell, the prince becomes a frog (warlocks trump). Today, a warlock and a witch each cast a death spell, and the

[26] Perhaps I should point out that in ordinary *preemption* cases there is a background circumstance at the time of the cause such that, in conjunction with either the preempting event or an actual preempted one at that time, the effect is implied (this fact implies the similar, but more oft-noted, fact that not only preempting but also *preempted* actual events belong to minimal sufficient conditions for the effect). Take, for example, the preemption depicted in Figure 3.1. Say $a$ and $d$ occur at time $t$. At $t$, certain neurons are properly positioned in an appropriate environment, and many events with the power to prevent $b$, $f$ or $e$ do not occur (e.g., meteorites, inhibitory firings other than $a$)—these $t$-circumstances do not lawfully imply $e$, but in conjunction with (the occurrence of) $a$ they do, and in conjunction with $d$ they do. In conjunction with $d$ they do because the only lawful way that $f$ could fail to occur after $d$ occurs (given we pack enough nonoccurrences into the $t$-circumstances just described) is for $a$ to occur—$a$ is the only event left that can prevent $f$. *But*, my point in the main text is an *intuitive* one, with which I am hoping you can sympathize: For "Rumpelstiltskin II," once I attend to the symmetry of the lawful implication, intuitively the example looks a lot like standard non-preemptive overdetermination, in that it looks as though the alleged cause is not really helping under the circumstances. This is in contrast to ordinary preemption: once aware of the lawful symmetry as it exists in Figure 3.1, it still seems there is a strong sense in which $a$ is helping under the circumstances.

prince dies. Call these two spells *war* and *wit*, respectively.[27] Let's say spells act directly; there are no intermediate stages whatsoever. This is a variant of a spell example that Schaffer (2000b) gives to illustrate "trumping preemption."[28] He would claim that *war* caused the prince's death and *wit* did not. The death does not depend on *war*, since without *war*, *wit* would be the highest-ranking spell and the death would still occur. So the example threatens to be a case of preemption involving no failure in the backup process.

As with "Rumpelstiltskin II," it is tempting, I think, to view *war* as somehow blocking or "deactivating" *wit*. But it is Schaffer's stated intention that there need be no such failure in or between a trumped event and the effect. A "neuron" example may make it a little easier to resist this temptation.[29]

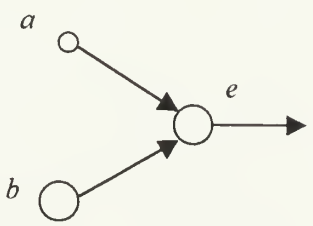


Figure 3.3

[27] To be clear: *war* is essentially a warlock spell, and *wit* is essentially a witch spell. I remain open as to whether each is essentially a *death* spell—either way is fine.

[28] His spell example is one in which the *earliest* spell of the day matches the prince's midnight fate. I forego this example for now because I think the use of a temporal trumping property makes things intuitively less clear and raises some special issues. (Anyway, it is clear that Schaffer does not want to confine trumping preemption to cases of temporal trumping properties.)

[29] The following example is a variant of one from Ned Hall (not in print). Schaffer also gives non-spell examples. One involves fields; but I think it is easy to be tempted to view a field as blocking or neutralizing another. Other examples involve human understanding and reacting—no crispness there, in my view. So I opt for this "neuron" example.

Suppose neurons fire in colors, and they fire in the color of the smallest neuron to fire upon them. *a* and *b* are both white firings, so *e* is a white firing. Here there is an intuition that only *a*, the firing of the small neuron, caused *e*—as before, preemption without failure in the backup process.

Say *a* and *b* occur at $t_1$, and *e* occurs at $t_2$. Let C be a thorough characterization of the background circumstances of *a* and *b* at $t_1$: the presence of the well-positioned *e* neuron in a suitable wet environment, the absence of any potential inhibitors of *e* (meteorites, etc.), the nonoccurrence of non-white firings upon the *e* neuron, the nonoccurrence of neurons non-identical to the *a* and *b* neurons firing upon the *e* neuron.[30] C does not lawfully imply *e*. But C plus *a* does, and, less obviously, C plus *b* does (the occurrences of the white *a* and *b* firings are the only $t_1$ firings consistent with C; so given C plus *b*, the death will occur: if *a* does not also occur, *b* is smallest, and if *a* does occur, *a* is smallest, yielding *e* either way). Thus, under the circumstances, *e* is *actually* implied by *b* as well as by *a*. This fact is *not ordinarily apparent*, but when I look at Figure 3.3 and attend to it, the case looks to me a lot like an overdetermination case, and it seems acceptable to count both or neither of *a* and *b* as causes. Again, the motivation for saying "neither" is the same as in standard non-preemptive overdetermination cases generally: each overdeterminer is occurring in circumstances that are already "enough" for the effect (C plus the other firing), so that neither is, under the circumstances, "helping." (It is crucial—so let me emphasize—that the point here is *not* that C plus *b* would be enough for *e* even without *a*. Rather, it is that C plus *b* *is* enough in the *actual* situation.)

[30] I am assuming that the neurons, and their firings, are essentially the size that they are (C would merely be more cumbersome to state, otherwise).

The point carries over to the spell case. Let C represent a thorough characterization of the background circumstances of the two spells: the presence of a live prince, the nonoccurrence of all spells that are not death spells, and the nonoccurrence of other witch and warlock death spells (and the nonoccurrence of spells of any other rank besides witch and warlock, if there be such a thing). C does not imply the death. But C plus *war* does, and, less obviously, C plus *wit* does (the occurrences of the *war* and the *wit* death spells are the only spell occurrences consistent with C; if at least one occurs, a death spell is highest-ranked). Thus, under the circumstances, the death is actually implied by *wit* as well as by *war*. When I attend to this hidden fact, the case looks to me a lot like an overdetermination case, and it seems acceptable to count both or neither spell as a cause. In particular, when I see that *war*'s circumstances *already* imply the death, it seems that *war* is not adding anything, not "helping."

As with "Rumpelstiltskin II," in a moral context, the overdetermination is more apparent. Imagine that Princess A and Princess B want the prince dead. They know that C will obtain. Each is ignorant of the other's plans. Princess A orders the warlock to cast *war*, Princess B orders the witch to cast *wit*, and the orders are carried out. As each knew would happen, the prince dies. It seems to me that, given the princesses' knowledge that C would obtain, the king would convict them both of murder; but if the witch had refrained because she witnessed A's order to the warlock (ordinary preemption), the king would convict B only of attempted murder. I am sure I would do as the king here. Perhaps the king and I are in error; but I submit that it is not at all clear that we are.

A family of examples involving ranking commands and orders is intuitively more compelling (for many) than the magical examples. Schaffer gives this one: the major

and the (lower-ranked) sergeant both order the corporal to charge, and the corporal decides to charge. We know that good soldiers always obey the highest-ranking order. It seems it was the major's order that caused the corporal's decision, and not the sergeant's.[31]

Knowing that soldiers obey the highest-ranking order, we are drawn to think that the corporal obeyed the highest-ranking order and not the other order. The corporal heard, understood, and acted in accordance with *both* orders; thus, in thinking he obeyed only the major's order, we are thinking of obeying an order as involving more than hearing, understanding, and acting in accordance with it. Indeed, as I will illustrate, it seems the notion of obeying, or complying with, or following an order has causal connotations, at least pragmatically if not literally, so that we are inclined to hear the relevant law—that soldiers obey the highest-ranking order—in a way that involves causal language. That is, the law (pragmatically) implies a causal law, which in turn begs the question of what are the causal facts of the case. If just before one of my peers leaves to eat lunch I deign to command him to eat lunch, and he understands me, and then he eats lunch, he may scoff at the notion that he was following orders, or obeying me: "I was going to eat anyway!" The claim that he was obeying my order sounds as if (even if it does not literally imply that) my order was a cause of his eating, that he ate because I told him to—not just that he acted in accordance with the content of the order. To say that soldiers obey (follow, comply with) highest-ranking orders connotes that highest-ranking orders *cause* soldiers to act, and probably this is a reason why the realistic command examples have more force than the magical examples.

---

[31] Schaffer credits Bas van Fraassen with this particular example (p. 175).

(Schaffer does not use "obey" or anything like it—he avoids explicitly stating the operative law, relying on our background knowledge of the military instead. But it is precisely in such terms that our background knowledge exists.)

Despite this, in a moral setting, even the realistic command examples look a lot like standard non-preemptive overdetermination. For the *fact* is, if both a major and a sergeant order a unit to burn a village and kill all the civilians, they will both be court-marshaled and (hopefully) convicted of war crimes—even if both orders are given and heard together by obedient soldiers. By contrast, if the major's order is transmitted in a way that interrupts the sergeant's order so that the sergeant's order is never received (ordinary preemption)—and this is known by the judges—then the sergeant will not be convicted of the slaughter (only of the illegal order, or the attempt to slaughter).

Schaffer points up that causal relations are often associated with other relations: causes tend to explain (via Deductive-Nomological argument[32]), to be evidence of and to be means to their effects; causes and effects are subsumed under the laws (see previous footnote); differences in the cause counterfactually imply differences in the effect (were neuron-firing *a* blue, *e* would be blue, while this is not true of *b*). He argues that sometimes all five of these relations hold between the trumping event and the effect while none hold between the trumped event and the effect. I will not take up all these arguments in detail. I accept the counterfactual implication point. I think the means argument he gives is incorrect, as I will explain below. The D-N explanation, evidence

---

[32] A D-N argument is one in which, using a law as a premise and also a non-law fact subsumed by the antecedent of that law, a conclusion is deduced: (P1) If x is a smallest firing upon a neuron y, then y fires in the color of x; (P2) *a* is a smallest firing upon neuron N; (C) N fires in the color of *a*. In this way, *a* explains why N fired white. (By the way, I do not think philosophers usually point to *D-N* explanation in particular when they note that causes explain their effects. But this is the sort of explanation Schaffer is pointing up here.)

and subsumption asymmetries depend on a wholesale exclusion of derivative laws—clearly, given C, both trumper and trumped can play these three roles, if derivative laws are used.[33] (Schaffer (p. 167) decries "tricky" laws, but we cannot simply label all derivative laws "tricks.") There is no question, for instance, that derivative laws can be used to collect evidence: given C in the neuron example, *b* is evidence of *e*, derivative law or not. That said, there is a partial asymmetry with respect to evidence. As Schaffer points up, usually the more we know about the cause, the more we can predict about the effect. Let D be the absence of other *a*-sized and smaller firings. Given D, the fact that an *a*-sized firing (*a*) occurred is evidence that the third neuron fired but not evidence that it fired white, while the further fact that *a* is white *is* evidence that the third fired white. By contrast, in *b*'s case I can find no background circumstance (counterpart to D) to support this sort of evidentiary relation of scale; C will not do it, for given C, the fact that any firing upon the third neuron occurred at all is evidence not only that *e* occurred but that *e* is white.

In sum, I do not think the asymmetry is as strong as Schaffer makes it out to be, and not strong enough to compel me to count one a cause and not the other.

The means argument is based upon Schaffer's original trumping example. In that example, the midnight fate of a prince matches the content of the *first* spell cast on that day; one day, Merlin casts a frog spell and, later, Morgana also casts a frog spell; there is an intuition that Merlin's spell, and not Morgana's, causes the prince to become a frog.

---

[33] The description of C suggests the relevant derivative law. In the spell case, we can have a very simple relevant derivative law on the assumptions that there are no ties for highest-ranked spell and that there cannot be infinitely many spells cast upon the prince: any spell not outranked by a different kind of spell matches the prince's fate. Given this law, each spell is subsumed and explains and is evidence for the effect.

The claim is that Merlin's spell, but not Morgana's, is an "effective means" to the prince's fate. Schaffer spells this out in the way Peter Menzies and Huw Price (1993) do, using what they call "agent probability": *c* is an effective means to *e* iff the probability that *e* occurs given that *c* occurs, where *c* is something an agent freely does or realizes, is greater than the probability that *e* occurs given that *c* fails to occur, where the failure is something an agent freely does or realizes.[34] (The point of *c* or its absence being freely brought about is that it then has its own independent (and very brief) causal history; this is supposed to keep it from raising the probability of past events, such as its genuine causes, thereby avoiding counting effects as means to their causes (Menzies and Price 1993, p. 191).) The idea that being effective, in the probability-raising sense here, is sufficient for causation is refuted in the probabilistic case—failed potential causes (Section 3.4.1) illustrate that—but the example here is deterministic. To my knowledge, there are no deterministic counterexamples. If Merlin's spell can be viewed as this sort of probability-raising means to transforming the prince, that should weigh strongly in favor of counting it a cause.

But I think neither spell is a probability-raiser in the sense required: the circumstances of each of the two spells include the other spell, also a frog spell; the example assumes determinism, and the spells are causally independent, so for each spell, whether or not that spell is freely realized, the probability of the *other* spell is 1, and so also the probability of the effect is 1. Schaffer argues to the contrary. He says Merlin's spell is effective because Morgana's spell might not happen—"Morgana is assumed free, so she

[34] The probability is to be objective, though beyond that, Menzies and Price (1993, p. 190) leave somewhat open how to make the probabilistic notions precise. Schaffer says that for his purposes we may take the inequality to say that P(*e* occurs and *c* occurs) / P(*c* occurs) is greater than P(*e* occurs and *c* does not occur) / P(*c* does not occur), where, again, it is assumed that *c*'s occurrence and nonoccurrence are freely realized.

might not cast her spell"—while Morgana's spell is *not* effective because Merlin's definitely will (did) happen (p. 169). But a mistake is made here in counting Morgana's spell as freely realized when the effectiveness of Merlin's is being tested. In using agent probability to assess the effectiveness of an event, we are not to take as freely realized *other* events we might wish to assess. In a deterministic world, it would only be *c* that we suppose to be freely brought about in assessing its effectiveness as a means to *e*.[35]

One last issue. There is a kind of probabilistic trumping that is interestingly different from the deterministic (Schaffer 2001, p. 80). Suppose warlock spells trump witch spells, in that, if both are cast (at any time on a given day), the type *and probability* of the prince's transformation are given by the warlock spell. Suppose warlock spells confer a .6 chance of transformation, witch spells a .9 chance. So, for example, if a warlock frog spell and a witch spell are cast, there is a .6 chance of a frog transformation and no chance of any other transformation, regardless of what type of spell the witch cast. One day, a warlock frog spell and a witch frog spell are cast on the prince, and the prince becomes a frog. The warlock spell may seem like the cause, even though the chance of the transformation would have been .9 without it, rather than .6.

Under the circumstances, the warlock's spell acts as an inhibitor: it reduces the chance from .9 to .6. The warlock may well have cast his spell in an effort to prevent the

---

[35] Would it help to have Merlin be the caster of each spell? No. While assuming that his first spell is freely brought about by him, we would still have to assume that all else is determined, *including his other actions*. (The "agent" in agent probability is supposed to be arbitrary: the question is whether an arbitrary agent's freely bringing about the first spell would make the prince's transformation more probable—in a deterministic world in which the second spell occurs—than would bringing about its absence. Clearly it would not. Having characters in the story play the arbitrary agents is slippery business: we must not simply make them free agents through and through, since the story is deterministic; only with respect to the realization of one action can they be assumed free. And remember that the two spells are causally independent, so the first's being freely realized cannot imply that the second is undetermined.)

transformation. Similarly, in a chancy version of the ordinary preemption situation depicted in Figure 3.1, *a* thwarts what may be the more reliable backup process, thereby reducing the chance of *e*—yet here we *do* count *a* as a cause. However, in this latter situation there are elements that seem to redeem *a*, such as the chain of probabilistic dependence via *b*—*a* helps an event that in turn helps *e*. There seems little, on the other hand, to redeem the warlock spell's damaging influence. In light of this, it is plausible to me that the warlock's spell is not a cause in these circumstances. I could go either way. I will not count it a cause.

# CHAPTER 4

# CONTEMPORARY COUNTERFACTUAL ANALYSES

## 4.1 Introduction

Here I describe most of the various contemporary counterfactual analyses of event causation, and I say why I think they are unsatisfactory. My major focus is on how well they handle the two central problems discussed in Chapter 3. I will take up deterministic analyses first (which need to solve only the preemption problem), then the broad analyses geared to handle all causation (these need to solve both problems).

Two recent deterministic analyses are very similar to the one I propose in Chapter 5: Christopher Hitchcock 2001 and Stephen Yablo 2002. I do not discuss these in the present chapter. I want to comment on how they relate to my proposal, so I will discuss them at the end of Chapter 5, after I have presented my analysis.

## 4.2 Ardon Lyon's Analysis

The first contemporary counterfactual analysis was offered by Ardon Lyon (1967). He was apparently unaware of the extent of the preemption problem and did not explicitly discuss it; rather, his main concern was that standard non-preemptive overdeterminers should count as causes.

Lyon had two requirements for causation. One was that the cause of *e* not be a "precondition" for *e* (p. 12). We need not discuss that, since it is the second requirement that preempting causes fail to meet: *c* is a cause of *e* only if at least one of the following holds (where *S* is some set of occurrent events preceding *e* and including at least *c* and one other, *b*): (1) If *c* had not occurred but all other events in *S* still had, then *e* would not have occurred; (2) if neither *b* nor *c* had occurred but all other events in *S* still had, *e*

would not have occurred, and if exactly one of *b* or *c* had not occurred but the other events in *S* still had, *e would* have occurred (pp. 8, 15–16).

Consider again the preemption problem. My dart-throw pops the balloon. Lucy was winding up to throw her dart, but she held back once she saw me throw. So without my throw, the pop would have occurred anyway, via Lucy. My throw preempted her windup, and consequently the pop is not dependent on my throw. The preemption diagram from Chapter 3 shows the essential structure of this kind of case (*a* is my throw, *d* is Lucy's windup, and *e* is the pop):

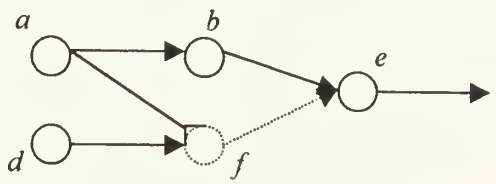


Figure 3.1

Now, for Lyon's clause (1) to be satisfied by *a* and *e*, there need to be (occurrent) events for *S* such that, were they still to occur, yet *a* not, then *e* would not occur. But there do not appear to be any such events. Let us consider (2), then. (2) was designed to count standard non-preemptive overdeterminers as causes, at least where there are only two overdetermining events: if neither overdeterminer had occurred, the effect would not have, but if just one had not, then the effect still would have (*S* can contain some other arbitrary events from the past). (2) succeeds in counting *a* as a cause of *e*, since *e* would not have occurred if neither *a* nor *d* had, and *e would* have occurred if just one of them had. Unfortunately, (2) also counts *d* as a cause, for the exact same reason. (2) fails to differentiate preemptive from non-preemptive overdetermination.

## 4.3 David Lewis's First Analysis and the Aftereffect Preemption Problem

David Lewis's first analysis (1973) identifies causation with the ancestral of what the deterministic simple analysis (DSA) identifies it with: *e* is *causally dependent* on *c* iff *c* and *e* are (essentially) distinct[1] and *e* is counterfactually dependent on *c*; *c* is a *cause* of *e* iff there is a finite sequence of occurrent events from *c* to *e* wherein each but the first is causally dependent on the one just before it (such a dependence chain Lewis calls a *causal chain* from *c* to *e*). As Lewis pointed up, this analysis allows a preempting event such as *a*, above, to count as a cause of the final effect, *e*: If *a* had not occurred, *b* would not have, and if *b* had not, *e* would not have. Here it is important to remember the proscription against backtracking (Section 2.2): we cannot say, "if *b* had not occurred, it would (have to) have been that *a* did not occur, hence *e* would still have occurred."

Remove *b* and the neuron it involves from the scene, and let there be laws of delayed action so that it seems *a* directly causes *e* at a (spatial and temporal) distance. Then there is no dependence chain from *a* to *e*. In his 1986d (pp. 202–3), Lewis addresses this case. He admits that his analysis' failure to count *a* as a cause of *e* is a shortcoming[2]—but he does not think it is a very important one, since it may be that our intuitions regarding it are not to be trusted. He seems to accept the possibility of such odd laws (he says, "that is possible, I take it"); but he thinks that our "common-sense judgements" about the causal facts may be wrong because we are not used to exercising causal judgements with action at a temporal distance in mind—such cases "violate the presuppositions of our habits of thought." I agree that the case at hand is of secondary importance, but I think it

---

[1] Lewis does not say whether he means *essentially* distinct, but it seems he does (see Section 3.2).

[2] "It would be better to agree with common sense about these cases, to be sure" (p. 203).

is so for a different reason. If we are going to count "action" at a distance as causation in simple scenarios not involving backup causes, then I think we should definitely count *a* as a cause of *e* in the described at-a-distance preemption situation. The only question, in my mind, is whether delayed action is really possible at all, despite seeming that it is. It takes a back seat for that reason.

In his 1986d, Lewis argues that, unfortunately, there is not always a causal chain, as defined, connecting preemptor and effect. He offers the following sort of diagram. Here, the upper process is two steps ahead, so that *d* and *e* are simultaneous.

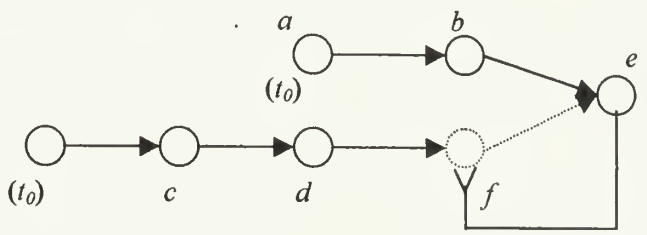


Figure 4.1

In this situation, the final effect itself stops the alternate process. This might happen if two rock-throwers were about to try to shatter a bottle and one thrower held back once he saw that the other had succeeded: the shattered bottle prevents the second throw. We see that *e* is not counterfactually dependent on *b*, since if *b* had not occurred, the lower process would have gone through and *e* would have occurred anyway, just a little later. So there is no causal chain from *a* to *e*.

This argument assumes that *e* is an event that (logically, metaphysically) *can* occur later than its actual time. Let me here interject some terminology that will be convenient from here on. Say an event is (temporally, spatially, etc.) *inflexible* if it could not occur

differently at all (with respect to time, place, etc.); say it is *flexible* if it could. An event's occurrence conditions or essence may be called flexible or inflexible as well, accordingly. Flexibility comes in degrees, whereas inflexibility is absolute.[3] The argument at issue assumes that some events are temporally flexible. A natural reaction to the argument is to deny this, to claim that if *b* were not to occur, the "*e*" that would occur would not really be *e*, but a different event. Lyon (1967, p. 8) glanced at this sort of case and had just this reaction to it; Lewis (1986d, p. 204) claims to have seen the matter this way when he wrote his 1973.

But Lewis later rejects this line, for two reasons (1986d, pp. 198–9, 204–5). First, we do not generally consider events to be temporally inflexible. We think that if the wedding had been delayed a minute, it would still have occurred. To the extent we wish to analyze our ordinary notion of event causation, it is preferable to stick to our ordinary views of events. Second, by insisting on temporal inflexibility it seems we will get spurious causes. Any event that affects the time of an event will come out a cause of it. Lewis gives the example of a poison victim's eating a big dinner before being poisoned. If the eating had not occurred, the poison would have taken hold earlier and the death would have occurred earlier. If the actual death is temporally inflexible, then it would not have occurred if the eating had not occurred. Yet, intuitively, the eating is not actually a cause of the death. For these reasons, Lewis and many others have preferred to accept that the type of scenario illustrated in Figure 4.1 constitutes preemption and probably refutes Lewis 1973.

---

[3] Lewis has a related term that often appears in the literature. He calls an event *fragile* "if, or to the extent that, it could not have occurred at a different time, or in a different manner" (1986d, p. 196). The built-in ambiguity of "if, or to the extent that" often calls for extra modifiers to be resolved; and the term has no obvious or generally employed antonym. So I will use my terms instead.

In fact, it appears the analysis already counts the eating a cause of the (temporally flexible) death, at least if we tweak the example just a bit. Let's just add that upon death, this poison will always dissolve away quickly (this doesn't alter the intuitive verdict that the dinner is not a cause of the death). Let *t* be some time before the actual time of the victim's death but after the time he would have died without the dinner—*t* falls within the delay period. Let *poisoned-man-at-t* be the event of the victim's presence in a poisoned state at *t*. If this event had not occurred, the death would not have occurred (because there would have been no poisoned man then or, as a result, thereafter); and poisoned-man-at-*t* is dependent on the dinner. When I see this causal path from the dinner to poisoned-man-at-*t* to the death, I am not sure it is wrong to count the dinner a cause of the death. So I think it is not clear that temporal inflexibility will give spurious causes. We could explore other examples, but I find none entirely compelling. The first of the above two motives is enough of a motive for me.

Preemption cases have to date been classified primarily along temporal lines, rather than by causal structure. The scenarios in the above two figures are universally referred to as "early preemption" and "late preemption," respectively, following Lewis (1986d, pp. 200ff). As Lewis originally cast the terms, they correspond to whether the failed backup process is cut off before the successful one has completed (where the final effect does not count as part of these two processes). Thus Lewis counts (alleged) preemption at a distance, as described above, as a kind of (alleged) late preemption. Throughout the literature, late preemption is supposed to be a kind of preemption in which there is no chain of causal dependence from cause to effect. That is the case in the at-a-distance example and in Figure 4.1. But not all cases of late preemption lack such a chain. Let

*the dinnertime ringing* be an event that can occur by way of a certain bell ringing within a few minutes of 6:00 p.m. At 5:00 p.m., you order Carl to ring the bell at 6:01, while I give a (single) order for Amy to ring it at 5:59 and Betty to remove and lock away the bell at 6:00 (regardless, I tell Betty, of whether the bell has been rung). Amy rings the bell at 5:59, Betty removes it at 6:00, and Carl then fails to hit the missing bell. It seems my order is a cause of the dinnertime ringing, since it causes someone to go ring the bell. If my order were not to occur, Carl would ring the bell, and *your* order would be a cause of the ringing. So my order preempts your order. The preemption is late, since the Carl process is cut off after the Amy process has completed (5:59). But there is a causal chain from my order to the ringing: If my order had not occurred, Amy's striking of the bell would not have occurred, and if this latter had not occurred, the dinnertime ringing would not have occurred, since Betty would have removed the bell at 6:00. Thus examples of late preemption display an interesting variety of causal structures, including dependence chains of the sort found in early preemption. As far as I can see, late preemption not involving delayed action makes trouble for Lewis 1973 only where the effect itself—or some effect of the effect—foils the backup process, as in Figure 4.1. In such examples, the disruption of the backup process is not simply after the effect but is an *aftereffect* of the effect at issue. In the dinnertime story, it is the effect-independence of the disruption that allows Lewis 1973 to escape trouble. If I instruct Betty to remove the bell *only if* Amy has rung it, then the disruption is an aftereffect, since the effect—the ringing—then contributes to Carl's failure, via Betty's removal; and the causal chain from my order to the ringing is thereby lost: If my order had not occurred, Amy's striking would not have occurred, but if Amy's striking had not occurred, Betty would not have removed the bell

and Carl would have rung it anyway. Let us call this *aftereffect preemption*. These are the specific late preemption cases that pose a threat to Lewis 1973.

At the end of Section 3.4.2.1, I said that preemption involves an incomplete backup process (Failure Thesis). This may seem doubtful, because the failed backup process preceding but not including the final effect need not be cut short at all, in that all of it may occur (Noordhof 1998; Paul 1998a). Take the case of two rock-throwers. As I told it, one potential thrower refrained when he saw that the bottle had been shattered. But we can imagine that both throws occur, and one rock just gets there first. In this case, the backup process is not cut short, save for the potential version of the effect that would have occurred had that process been successful (here we ignore the missing intermediate of rock-contacting-bottle; or, we can just make this the final effect at issue). Or in the dinner-bell example, if Betty mutes the bell instead of steals it, and if Carl strikes the muted bell, then the Carl process (preceding the potential 6:01 ringing) is complete.

But the Failure Thesis counts the final effect as part of the backup process. If we add the effect to the processes just described, then a potential *part* of the backup process fails to occur in the sense that the backup-process *version* of the effect does not occur: the late shattering, which would have occurred had the first throw not occurred, does not occur; and likewise the 6:01 version of the ringing does not occur. (Such a version, were it to occur, would be a non-proper part of the flexible effect at issue—see Section 2.4.2.)[4]

[4] There are other noteworthy failures, which perhaps an analyst could capitalize on. One process leading to the shattering is the persistence of the bottle. This persistence at the time just before the bottle would have shattered if the second rock had been the cause is a backup that is foiled by the first shattering (the bottle is shattering at that very time). In the bell case, the nonoccurrence of a muting hand on the bell is a backup cause foiled by Betty's muting hand, if we count nonoccurrences as causes.

## 4.4 Marshall Swain's Analysis

Marshall Swain (1978) modifies Lewis's analysis so as to be able to count standard non-preemptive overdeterminers as causes. But like Lyon's analysis, Swain's ends up counting as causes not only non-preemptive overdeterminers but also preempted backups, and thus it constitutes a step backward with respect to our main concerns here.[5] On Swain's analysis, a sufficient condition for *c* being a cause of *e* is the following (where "causally prior" is a certain complex world-theoretic relation designed by Swain to "guarantee" that causation is asymmetric—it is not relevant to preemption, nor does Swain, who is well aware of preemption, claim that it is):

> Some set of events $D = (d_1, d_2, \ldots, d_n)$ occurred (possibly having only one member) such that
> (a) if *c* had not occurred, and if any member $d_i$ of *D* had occurred, but no other members of *D* had occurred, and if *e* had occurred anyway, then there would have been a causal chain [in Lewis's sense] from $d_i$ to *e* consisting wholly of occurrent events, and $d_i$ would have been causally prior to *e*; and
> (b) if no member of *D* had occurred, and if *c* and *e* had occurred anyway, then there would have been a causal chain from *c* to *e* consisting wholly of occurrent events, and *c* would have been causally prior to *e*. (p. 16)

The preemption examples above refute the sufficiency of the condition. To be sure I am applying it correctly, I will just mimic one of Swain's own applications: he applies it to a simple overdetermination case involving three occurrent events; I will use his exact words, just substituting the names of three occurrent events from one of the above preemption scenarios. Consider either of the above diagrams, and let us see whether the condition yields that *d* is a cause of *e*. "To see that *d* is a cause of *e* we may note that there is a set of events *D*, consisting of the event *a* alone, such that if *d* had not occurred

[5] Swain was pursuing various other refinements; he considered the preemption problem solved by Lewis.

but *a* had occurred anyway, then there would have been a causal chain of occurrent events from *a* to *e* and *a* would have been causally prior to *e*. Moreover, if *a* had not occurred, but *d* had occurred anyway, then there would have been a causal chain of occurrent events from *d* to *e* and *d* would have been causally prior to *e*" (p. 16, with altered event names). This all appears to be true, so *d* comes out a cause of *e*. (Note that the requirement that there be a chain of *occurrent* events does not save the condition. For despite the fact that actually nonoccurrent events would occur between *d* and *e* in the antecedent situations invoked, a "causal chain" of occurrent events would exist because there would be a one-step dependence of *e* on *d*, and both of those are occurrent.)

### 4.5 Lewis's "Quasi-dependence" Solution for Aftereffect Preemption

Lewis (1986d) attempts to amend his analysis to accommodate aftereffect preemption. On the revised analysis, still an event *c* is a *cause* of an event *e* iff there is a causal chain from *c* to *e*. But now a *causal chain* is a finite sequence of occurrent events such that each one (except the first) is either causally dependent or "quasi-dependent" on the one just before it. Quasi-dependence is supposed to capture the intuition that "whether the process going on in a region is causal depends only on the intrinsic character of the process, and on the relevant laws"[6] (1986d, p. 205). Lewis never explicitly lays out a definition of quasi-dependence. I have culled one from what he does say:[7] Occurrent event *b* is *quasi-dependent* on occurrent event *a* iff *a* is first and *b* is last in some ordered

---

[6] I leave alone the question of whether this intuition is correct. (If standard non-preemptive overdeterminers are not causes, then surely it is not. Other cases are less clear, because they involve a law whose very relevance to the process depends partly on factors extrinsic to it. Consider a law that invokes circumstances; or consider perhaps the sorts of extrinsic laws involved in the examples of Sections 3.4.2.2.3 and 3.4.2.2.4 (the latter, "trumping," which Lewis takes seriously).)

[7] 1986d, pp. 176–7 and 205–6—especially bottom p. 205. I am aware (i)-(iii) may contain redundancies.

"process" *p* of events that is such that, with respect to the ordering, the great majority of lawfully possible processes intrinsically just like *p* (i) have their later parts dependent on their earlier ones (as Lewis puts it), (ii) have their last event dependent on their first, and (iii) are "causal, according to [the] original analysis." Lewis says that the "great majority" is to be measured by varieties of surroundings, though he does not say what constitutes "great" or how to carve up the possible surroundings into varieties. Of a "process" he says only that it is a "course of events" that may or may not be causal (this is not *my* use of "process" above, which has always meant causal process).

There are some puzzling things here. Does (i) mean that each event is dependent on the one just before it in the ordering? (iii) is puzzling because "causal process" is not an element of the original analysis. It *seems* that (i) and (iii) are calling for a *causal chain*, as originally defined. I think we can press on here without answers to these questions.

In the situation of Figure 4.1, arguably *e* is quasi-dependent on *b*, and *b* is causally (as well as quasi-) dependent on *a*, so *a* and *b* come out causes of *e*. Presumably *e* is supposed to be quasi-dependent on *b* because it is only in very special circumstances that a course of events just like *b*-*e* is accompanied by overdeterminers, of one sort or another, of *e*. *d* is not a cause, since *e* is neither quasi- nor causally dependent on it: it is not the case that in the great majority of varieties of circumstances where a course of events just like *d*-*e* occurs there is dependence, since *d* and *e* are simultaneous. (As for *c*, it is at a spatial distance from *e*, so plausibly there will not be dependence there in the great majority of varieties of circumstances.)

I do not think it is clear that this works. Quasi-dependence introduces vague and undefined terminology. "Intrinsic" and "course of events" are not too troubling to me—

we have a strong intuitive understanding of the first, and for the second I think Lewis could get by with *sequence* of events instead—rather, it is the idea of the great majority of varieties of surroundings: we do not know how to carve the possibilities into varieties, or how great "great" is. Thus I think it is not clear that *e* is *in general* quasi-dependent on *b*, even if it is in the case of neurons. Take the aftereffect example where two rocks are thrown at a bottle, one arriving first and shattering it. A late stage of the first throw is a (direct) cause of the shattering, but it is not clear that in the vast majority of varieties of surroundings for that precise <stage, shattering> process (sequence), there is no backup cause of the shattering around. For on the face of it, there are many varieties of event that can shatter a bottle—different types of projectile, but also many various sources of vibrations, shock waves, sudden heat, cold, and air-pressure change. (Such events could occur simultaneously with, or after, the shattering, so that the intrinsic nature of the causal sequence is as it actually is.)

That is a potential counterexample to the necessity of the chain of the disjunctive relation *dependent or quasi-dependent* for causation. In the literature, there are various alleged counterexamples to the sufficiency of quasi-dependence for causation. I find them all to be of doubtful success and shall not discuss them.[8] But I have another kind of apparent counterexample to the necessity of the disjunctive chain to offer, as follows.

---

[8] Jonardon Ganeri, et al. (1996); Douglas Ehring (1997); Ned Hall (1997); L. A. Paul (1998a). ( I think Ehring's example is promising but that it needs changes to be more convincing. The way I see to change it turns it into a trumping case, where the effect quasi-depends on a trumped event; as I argued (Section 3.4.2.2.4), it is not clearly wrong to count the trumped event a cause. The other three examples use uncharitable interpretations of "great majority of varieties" or of "intrinsically just like" and can, I believe, be resisted in a principled way.)

4.5.1 Meta-prevention Preemption

As the trapeze artist begins to fall, I commit murder by untying the rope holding up his safety net. The untying was among the causes of his death—a meta-preventing cause (it prevented a later presence of net from preventing the death; Section 3.3.5). But had I not done it, my partner Clyde would have cut the rope—as it was, he just stood and watched—so the death would have occurred anyway. As Figure 4.2 makes clear, there is no disjunctive chain of dependence or quasi-dependence from the untying to the death.

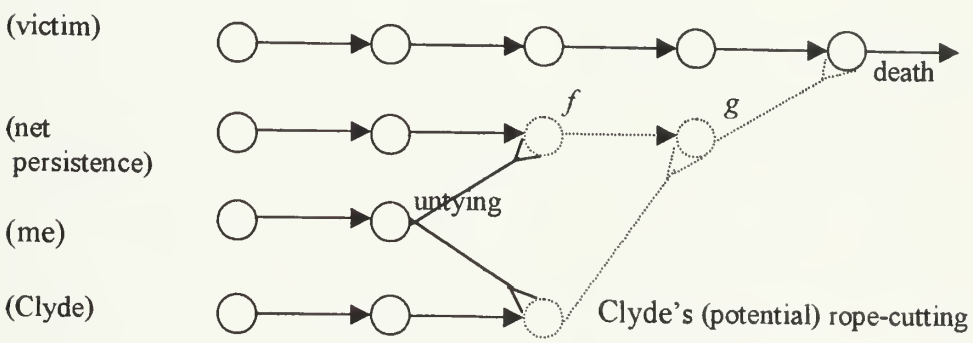


Figure 4.2

Here is a potential repair. We could allow "nonoccurrences" into the chain, such that for a nonoccurrence of *e* not to "occur" is (by stipulation) just for *e* to occur. Now notice that if the untying hadn't occurred, *f*'s nonoccurrence wouldn't have, and if *that* hadn't, the death wouldn't have. This way we can count the untying a cause of the death.[9] This is a decent idea. After all, as meta-prevention shows us, sometimes nonoccurrences seem to be causal intermediates, so it is fitting that a causal-chain approach should utilize them.

[9] Schaffer (2000c) suggests that every event is a pair <Property, Region>, where the property may be negative: e.g., <absence of firing, neuron region>. To suppose this event to occur is to suppose there is no firing in the neuron region; to suppose it *not* to occur is to suppose there is. Admitting such events may be another solution.

That example had nothing to do with aftereffect preemption. The message it sends to Lewis 1973 is that, since preemptors can be meta-preventers, you are going to have to include nonoccurrences in your chains. But when now we notice that meta-preventers can be *aftereffect* preemptors, we find that Lewis 1973 is still in hot water, and quasi-dependence is the wrong kind of tool for getting it out. Thus imagine there is a small earthquake; Rose tries to reach for the Ming vase on the table to prevent its fall, but I maliciously hold her back and prevent her from doing so; the vase falls. A second afterward, Rose's big dog bumps the table, and this would have caused the fall had the vase still been on the table. If we are serious about counting meta-preventers as causes, then my holding Rose should count as a cause of the fall despite the preemptive setting. Figure 4.3 shows that the causal structure here can be made precise and leaves no room for a quasi-dependence solution.

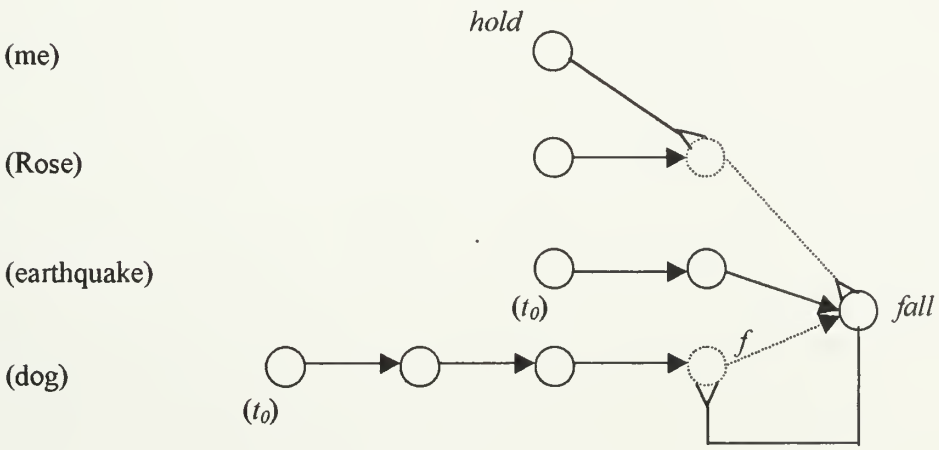


Figure 4.3

## 4.6 L. A. Paul's Solution for Aftereffect Preemption

L. A. Paul (1998b) proposes a simple solution to the aftereffect preemption problem: again identify causation with the ancestral of causal dependence, but redefine causal

dependence such that *e* depends causally on *c* iff, were *c* not to occur, then either *e* would not occur or *e* would occur later than the time it actually occurred. Looking at Figure 4.1, we see that if *a* had not occurred, *e* would have occurred later, while if *d* had not occurred, *e* would have occurred just when it actually did—*a* comes out a cause and *d* does not. Similarly for the example of Figure 4.3: without the hold, the fall would have occurred later, while without the dog's bump, the fall would have occurred when it actually did. The proposal will not handle the trapeze example (Figure 4.2), but as before we could allow nonoccurrence links in the chains.

The analysis has the result that to hasten an event is always to cause it. But there is no implication that events are temporally inflexible, or that they cannot occur later: an event is caused when it is hastened, but no claim is made that this is because a later effect would have been a different event.

Other writers (e.g., Jonathan Bennett (1987) and Penelope Mackie (1992)) have expressed sympathy for the idea that hastening is always causing. I know of no compelling case to be made either for or against it; I will not delve into that literature here. It seems to me not an unreasonable idea, if not particularly compelling. Anyway, there is yet another brand of preemption that Paul's proposal falls prey to, as follows.

## 4.7 The Process Preemption Problem

The 1000-meter fall was a cause of the man's death; the bowling ball's journey down the lane caused the pins' fall; the train journey caused the train arrival; the loud bang was caused by the door's swing shut: we regularly take events that are causal processes

leading to some final event as causes of that final event.[10] And the corresponding counterfactuals seem to support this: If the door's swing shut hadn't occurred, the bang wouldn't have occurred; without the fall, the death wouldn't have occurred.[11] This is no less so in preemption contexts: the beginning of the bowling ball's roll down the lane may prevent someone else from rolling *her* ball at the pins, but no less does the first ball's roll cause the pins' fall.

Consider Figure 3.1 again. It appears the composite event *a-b* is a cause of *e*. But DSA fails to count it as such. Clearly, if *a-b* were not to occur, *a* either would not or might not occur (nothing licenses us to say that if *a-b* were not to occur, *a would* still occur, any more than that if the 1000-meter fall were not to occur, all but the last bits still *would*[12]); but then if *a* might not occur, by the same token the backup process might go through and *e* might still occur. DSA suffers from the *process preemption problem*: A "process cause" of *e* fails to be counted as a result of its having a preemptor as a part.

---

[10] But doesn't this have punch: "Not the *whole* door-swing was a cause of the bang, only the last bit was"? I hear it, but it is too mysterious what is intended by this statement to give it any serious weight. Clearly other small or instantaneous parts besides the last (if there is a last) *were* causes, just as *a* is a cause of *e* in Figure 3.1, so it is clearly false that only the last bit was. Then what is the quoted statement trying to express? Perhaps the intent is "not all smallest parts were *direct* causes of the shattering, only the last one was." But it is too hard to say.

[11] If we try, we can begin to doubt these counterfactuals by invoking inflexible occurrence conditions for the process: if the door-swing cannot occur much differently, then it seems it might fail to occur by the door swinging a little slower, whereupon the (temporally flexible) bang still occurs. But let us not go out of our way to falsify what sounds true. A charitable and common-sense interpretation of the event-designators can apparently allow the counterfactuals to be true: just let the occurrence conditions have a certain moderate flexibility. An event fails to occur iff none of its possible versions occur, so if we take the door-swing to be moderately flexible, its nonoccurrence will imply that nothing moderately similar to the actual swing occurs; and we would accept that had nothing at least moderately similar to it occurred, then there would have been no bang.

[12] Our rejection of these claims suggests the need for a revision of Lewis's view that the antecedent-worlds closest to actuality are ones in which the past matches the actual past as much as possible consistent with the truth of the antecedent and a smooth transition (1979). For apparently, we do not hold the past fixed *into the time of the antecedent event itself*, even when doing so would maximize past match.

Lewis 1973 also falls prey, since there are no causally intermediate events to generate a chain from *a-b* to *e*. And so does Paul's revision, since her temporal disjunct is of no use here: if *a-b* were not to occur, *e* might occur just when it actually did. (And clearly the addition of nonoccurrence links is no help.)

Now if we permit quasi-dependence links in a causal chain, perhaps process preemption poses no special difficulty—perhaps there is quasi-dependence of effect on process. But now we are stuck with a less clear, and not clearly successful, analysis, due to the notion of quasi-dependence (Section 4.5), and we are stuck with a fairly clunky analysis: a link in the chain can be one of counterfactual dependence (maybe involving a nonoccurrence), or quasi-dependence, or that the one event would have occurred later without the other. It would be better to have something more unified in form and in motivation, and something more clearly adequate.

One might reply that it is fair for analyses to be restricted to atomic, or at least instantaneous, events, as we may start with that as a first step and build broader analyses from there. Perhaps, but not all analyses need to duck composites. Many of the analyses below (including my own, in the next chapter) do not try to solve the preemption problem by assuming that preempting causes are always indirect; they do this by focusing on the failure of the backup process, which is independent of whether the preemptor is indirect; consequently process preemption poses no special difficulty for them. If we do not need to address composites as a separate step, all the better. Doing so would add complexity; also, it does not look easy, and no one has proposed how it might be done. Nor is it just a minor addendum to causal analysis, for the actual events of which we know and talk tend to be extended, and we often think of them as extended (I do not see that we generally

idealize large events as atomic or instantaneous, though perhaps sometimes we do); an analysis for instantaneous or atomic events would appear to have little application on its own.

## 4.8 The "PCA*" Analysis

Jonardon Ganeri, Paul Noordhof and Murali Ramachandran (1998) revise their original "PCA" analysis (1996) in response to counterexamples by Alex Byrne and Ned Hall (1998). The revised analysis—dubbed "PCA*" by its authors—is given as follows (the events in the two definitions need not be occurrent (actual)).

*Def 1.* For any events $x$ and $y$, and any set of events $\Sigma$, $y$ $\Sigma$-depends on $x$ iff (i) if neither $x$ nor any of the events in $\Sigma$ were to occur, then $y$ would not occur, and (ii) if $x$ were to occur without any of the events in $\Sigma$, then $y$ might occur.[13]

*Def 2.* For any events $x$ and $y$, and any set of events $\Sigma$, $x$ is a $\Sigma$-ancestor of $y$ iff there is a chain of events, $z_1, \ldots, z_n$, such that $z_1$ $\Sigma$-depends on $x$, …, and $y$ $\Sigma$-depends on $z_n$.

(PCA*) For any actual, distinct events $c$ and $e$, $c$ causes $e$ iff
there is a (possibly empty) set of possible events $\Sigma$
such that
(i) $c$ is a $\Sigma$-ancestor of $e$, and
(ii) every $\Sigma$-ancestor of $e$ is an actual event.

The central intuition is that a sequence forms a causal chain when, had certain extraneous events (still) not occurred (certain backup causes, for instance), each element would have been counterfactually dependent on the one before and done so without the mediation of

[13] "Might" here will do no noticeable work over "would." Ganeri, et al. use it to deal with what they call "'accidental' causation." Imagine we take it to be a "fluke" or "accident" that someone's action $c$ succeeds in causing $e$, so that, even though the situation is deterministic, our general feeling is that $c$ might have failed to cause $e$. The authors want to allow that if things had been a little different in such-and-such a way causally irrelevant to $c$ and $e$, and yet still $c$ had occurred, $e$ might not have occurred (p. 221). I am not sure I agree, but I do not see anything that hangs on it.

any non-actual events. "Speaking loosely, we may say causes are 'potentially complete ancestors' of their effects" (p. 222).

Where $\Sigma = \varnothing$, $\Sigma$-dependence between actual events reduces to clause (i) of Def. 1: If $x$ and all the events in $\varnothing$ were not to occur, then $y$ would not occur. We see that among actual events, a chain of causal dependence, in Lewis's sense, implies a chain of $\Sigma$-dependence, with $\Sigma$ empty. Thus in Figure 3.1, $a$ is a $\varnothing$-ancestor of $e$. And all $\varnothing$-ancestors of $e$ are actual: no actual event can be $\varnothing$-*dependent* on a non-actual one, so no non-actual event can be a $\varnothing$-ancestor of an actual one. Thus a chain of causal dependence implies causation on PCA*.

This is no help with aftereffect or process preemption, but that is because in identifying $\Sigma$ with $\varnothing$ we are not yet taking advantage of the machinery. Let $\Sigma$ be $\{f\}$. For either Figure 3.1 or 4.1, if $a$ were not to occur and (still) $f$ were not to occur, then $e$ would not occur, so $a$ is a $\Sigma$-ancestor of $e$; and every $\Sigma$-ancestor of $e$ is an actual event, because supposing away some non-actual event in addition to $f$ leaves us at the actual world and can thus never generate a $\Sigma$-dependence step. This all holds for $a$-$b$ as well as for $a$, so PCA* handles process preemption. Do spurious causes arise in these scenarios? No. We need to put $a$ or $b$ in $\Sigma$ in order to get $e$ to be $\Sigma$-dependent on an event in the backup process. But then not every event on which $e$ is $\Sigma$-dependent will be actual, for it will be $\Sigma$-dependent on $f$: If $f$ and every member of {b} had not occurred, $e$ would not have.

Consider the sort of aftereffect case in which a failed backup process preceding the final effect is not cut short—e.g., two rocks are thrown at a bottle, one gets there first and

does the shattering. Ganeri, et al. consider a similar example (1996, p. 221). Let $r_2$ be a stage of the second rock's journey. Let us see whether $r_2$ comes out a cause. Let $\Sigma$ contain a stage, $r_1$, of the first rock's journey. The shattering is $\Sigma$-dependent on $r_2$, satisfying clause (i) of PCA*. How about clause (ii)? Ganeri, et al. say that the shattering is $\Sigma$-dependent on the non-actual event of the second rock contacting the bottle: without that contact and without $r_1$, the shattering would not occur. True, but as Paul (1998a) points out, there appears to be trouble if we back up a little. That is, let rock-contact-with-bottle be the effect at issue (where this is an event that can involve either rock). *Then* what non-actual event will be a $\Sigma$-ancestor of the effect? Paul says none. Well, arguably there are two, as we have already seen, so there is no real trouble after all. For one, there is the presence of intact bottle at the actual time of the shattering. If $r_1$ and this presence had not occurred, no rock-contact would have occurred. Second, there is the essentially late, temporally inflexible rock-contact that would have occurred had the first rock not been thrown. If $r_1$ and this late contact had not occurred, there would have been no contact. So there is, after all, a non-actual event that is a $\Sigma$-ancestor of the contact.[14]

PCA* is quite effective, then. And there is a bonus: at-a-distance preemption is handled. In a revised Figure 3.1, where $b$ is removed so that $a$ directly causes $e$ at a distance, let $\Sigma$ be $\{f\}$ as before and $a$ comes out a cause of $e$ according to the same reasoning as that two paragraphs up. Nothing hangs on intermediates in these examples.

[14] This same response applies to Byrne and Hall's time-delay version of aftereffect preemption (1998, footnote 6, p. 44).

One can easily see that PCA* counts standard non-preemptive overdeterminers as causes. Some may take this as a good result. It counts trumped events and $s_2$ in "Rumpelstiltskin II" as causes (let $\Sigma$ contain the trumping event in the first case, $s_1$ in the second)—some may take this as a bad result, though I have argued that it is not clearly wrong (Section 3.4.2.2).

The result I do not want to accept is that for allegedly preemptive prevention (Section 3.4.2.2.1) PCA* counts the backup disrupter as a cause (Byrne and Hall 1998). We had the example from McDermott where I catch a ball that is headed toward a window, and between me and the window is a brick wall. I do not want to count the presence of the wall as a cause of the presence of intact glass. But let $\Sigma$ contain my lunge for the ball; the presence of the brick wall is a $\Sigma$-ancestor of the intact glass, and no non-actual event is.

## 4.9 Lewis's "Influence" Analysis

Lewis (2000) recently produced a new and very different counterfactual analysis of event causation. It goes as follows.

A *fragile* event is one with little flexibility—it cannot vary much with respect to time, place or manner. An *alteration* of event *e* is "either a very fragile version of *e* or else a very fragile alternative event that is similar to *e*, but numerically different from *e*" (p. 188). Lewis does not want to worry about whether these alterations are numerically different events from *e* or just different versions of *e*: "We ought to make sure this distinction bears no weight in our analysis." Note that if *e* occurs, an alteration of *e* occurs—just one, in fact, according to Lewis (suggesting that these alterations are *absolutely* fragile, inflexible in every respect). Lewis writes, "where *c* and *e* are distinct actual events, let us say that *c influences e* if and only if there is a substantial range $c_1$, $c_2$,

... of different not-too-distant alterations of $c$ (including the actual alteration of $c$) and there is a range $e_1$, $e_2$, ... of alterations of $e$, at least some of which differ, such that if $c_1$ had occurred, $e_1$ would have occurred, and if $c_2$ had occurred, $e_2$ would have occurred, and so on." *Causation* is defined as the ancestral of influence (pp. 190–1).

A few expository comments are in order. The correspondence between the $c$'s and the $e$'s need not be one-one and can be as askew as many-two. This is a result of the phrase "at least some of which differ," which permits that there be only two $e$ alterations named by the "$e_i$"s. Thus imagine a prince susceptible only to hog spells cast in exactly the right fashion. A hog spell is cast, and the prince becomes a hog. Of course, were the actual cause alteration to occur, the actual effect alteration would occur—there is one mapping. Were the cause at all different, then in place of the actual effect alteration would be a different one, a standing man; so all the close, non-actual cause alterations map to this one non-actual effect alteration. In this way, such all-or-nothing effects can still be counted.

Lewis: "Influence admits of degree in a rough and multidimensional way.... Plainly, there are many ways in which something can be more of a cause of some effect than something else is, even if it is not an all-or-nothing difference of influence versus no influence" (pp. 190–1). Lewis indicates that degree and way of influence are sensitive to how numerous and varied the $c$'s are, how distant (dissimilar) they are from each other, how distant they are from the actual alteration of $c$ (this one is especially important, he says), and how much the $e$'s differ from one another. Perhaps there are other factors, as well; he seems not to rule it out.

The requirement that the range of *c*'s be "substantial" appears to be without consequence, for if *some* close *c* range *R* maps to multiple *e*'s, it seems the *entire* superset *S* of close *c*'s will, too. The *S* members will map either just to those same *e*'s or to those plus some others. This will be false only if the members of *S–R* map to things that are not *alterations* of *e* at all. In light of Lewis's general discussion and examples, I cannot imagine such a case; Lewis never offers one. But even if "substantial" makes no difference to whether one event counts as a cause of another, it may still have a point. As just explained, the degree and way of influence are sensitive to how *c*-alterations map to *e*-alterations; the analysis tells us to look at a substantial range of *c*'s when assessing the nature of a mapping.

Gravitational effects abound. We do not want to count my driving toward the beach as a cause of the local high tide...well, not unless we are, in fact, concerned with extremely miniscule aspects of the tide. Lewis's idea is that, in our use of "cause," pragmatically we ignore small amounts of influence that fall outside the range we are interested in at the time—just as whether we count a vessel empty depends on what sorts of things we are ignoring (often air, a little dust) (pp. 188ff).

Lewis writes that in supposing away an event, we usually imagine it to be "completely and cleanly excised from history, leaving behind *no* fragment or *approximation* of itself," (p. 190, my italics). He also says, "alterations may include some in which [the event] is completely excised," and he explicitly allows that some such alterations count as "not too distant" from the actual alteration.[15] Thus, what is in the region of an event after we

[15] Lewis 2000, p. 190, second full paragraph. I read it that "these alterations" (fourth sentence) are the ones we "could look at" and the ones he is trying to capture in the analysis, later in that paragraph.

suppose it not to occur may count as an alteration not too far from the original while being no approximation to the original. Apparently, then, there are two similarity relations at work here. I think the "no approximation" is with respect to how we would intuitively compare two regions for intrinsic similarity, "off-the-cuff" and without any special interests; the closeness of alterations is something different, which presumably we are attuned to in causal contexts. Lewis does not discuss this second similarity relation, and I am not sure I really have a sense of it.

In Section 3.4.2.2.4, I discussed an alleged kind of preemption, "trumping preemption," and said that I was not convinced it was truly preemption. Lewis says his analysis rightly takes it as genuine. He considers the example in which the major's order trumps the sergeant's and they simultaneously order the troops to advance. Different orders from the major would have been followed by different actions by the troops, while this is not so for the sergeant; so it appears the analysis counts the major's order a cause and not the other.

But it seems to me that if either order had occurred earlier, the obedient troops would have acted earlier—and if the earlier order had also been different, the troops would have acted earlier and differently—so it appears that both orders come out causes. Lewis does, mind you, explicitly allow that temporal alterations may count as close, and he insists that there is nothing special about temporal alterations (p. 187). We can imagine that we know that the sergeant and the major habitually try to find the opportune time to issue an order; then, intuitively, an earlier order will not seem far-fetched at all (while still, I think, it will seem that, if the orders are given simultaneously, only the major's order caused the troops to advance).

Lewis could fairly say that the analysis correctly counts the major's order *more of* a cause than the sergeant's: close alterations of the major's map to a *wider array* of effect alterations than do close alterations of the sergeant's, since the alterations to which they map include present ones as well as earlier ones; second, present alterations are generally *closer* to the actual alteration than past ones. Still, the mapping for the sergeant's order is fairly rich, and I had thought—we were told—that the goal was to exclude it as a cause. Perhaps the claim will be that it is only a cause in a subtle and special *way* that we are not interested in when we first hear the example—a way quite different from the way in which the major's order is a cause. But I do not see what this is: judging from the alteration mappings, it seems both orders influence the time and manner of the effect. Perhaps, despite Lewis's claims, temporal alterations *are* special somehow, and somehow we should discount the earlier alterations of the sergeant's order.

By going through the diagrams, one can easily see that the analysis appears to count preempting causes: there is always a chain of influence from cause to effect, sometimes via a nonoccurrence—or "absence," as Lewis prefers—sometimes in virtue of influencing the time the effect occurs (absence counts as a kind of alteration, according to Lewis (p. 190)).

I do not think that "influence" suffices for causation. Consider the Merlin/Morgana trumping example (Section 3.4.2.2.4). Suppose that within range of Merlin's and Morgana's frog spells was a second prince, but that he was immune to frog spells. So at midnight this second prince remained human, whistling as he strolled through the garden. This second prince is sensitive to other kinds of spell, so had Merlin cast a different one, the whistling would have been absent; yet it seems wrong to me that Merlin's actual spell

should therefore count as a cause of the prince's whistling. If it were true that, had Merlin's frog spell not been cast, a different spell would have been cast instead, *then* it would be true that his spell was a cause of the whistling, because then it would be true that if it had not occurred, the whistling would not have occurred. But we can just stipulate that there are no special circumstances to make that true—instead, any variety of spells or none might have occurred if Merlin's frog spell had not.[16]

Here is a similar, this-worldly example. Suppose two rocks are thrown at a window that has a hole in it a little bigger than the rocks. One rock arrives first and passes easily through the hole, not touching the glass at all; the second rock hits the glass and shatters it. The second throw—by which I mean here the windup and release, excluding the rest of the rock's journey—was a cause of the shattering, and it seems the first throw was not. After all, we want to say that the first throw *almost* caused the shattering, because that rock almost hit the glass (first); we don't want to say that it *did* cause it. But it may be that some, or many, of the slightest differences in the first throw would have been magnified as that rock moved through the turbulent air such that it would have missed the hole and shattered the window, in which case it "influenced" the shattering. Notice that this could be true in an aftereffect preemption version, too: it may be that some, or many, slight alterations in the second release would have allowed the second rock to overtake the first before they reached the glass.[17]

---

[16] Schaffer (2001a) makes a similar objection.

[17] Notice that the nature of the alteration mapping in these examples is similar to that between trumping event and effect. But you did not think of the trumping event as a cause in this slipping-through-the-hole way, did you? (After all, trumpers make intuitively more compelling candidates as causes than do the alleged causes in these other examples.) To the extent the analysis counts trumpers as causes, still it seems not to count it in the right way.

My second complaint is that Lewis's analysis cannot handle complex events. Let us first look at spatiotemporally composite events: The conference lasted four days. On day 4, a tiger's entrance through the window influenced the conference; the conference influenced the newspaper headlines on day 2, since on day 1 several shocking conference papers were delivered. So by a two-step chain, the tiger's entrance on day 4 comes out a cause of the newspaper headlines on day 2. In this way, with enough steps (or just large, scattered events), we can get almost any event to be a cause of any other.

This flaw is a step backward, in that Lewis 1973 does not have this problem. Let us suppose that the difference the tiger's entrance made to the conference is that there was a commotion on day 4 in place of a certain paper being delivered. Clearly, conferences that either essentially have this commotion or essentially lack that paper are (counterfactually) dependent on the tiger's entrance. Any conference that only *accidentally* has this commotion and paper-lack is *not* dependent on the tiger's entrance, since without the entrance, nothing essential would be missing. So *only* conferences that essentially have this commotion or paper-lack are dependent on the tiger entrance. But the headlines on day 2 are dependent on *no* such conference, since such a conference *might* fail to occur simply in virtue of a lack of commotion or a paper delivery (as these things are essential), thus leaving room for the shocking papers still to occur on day 1 (they *might* still have occurred)—hence no conference is a middle link in a chain of dependence from the tiger's entrance to the headlines. Lewis 1973 is safe.

In that example, the tiger entrance affects one part of the conference, but a different part affects the headlines. Likewise, there are cases in which the intermediate event is an effect by way of one feature, or aspect, and a cause by way of another. Jig-chuen Lee

(1988) proposes a counterexample to causal transitivity generated from such an example. Here is a rendition of it. Someone pours green food coloring into the poisoned drink. Blind Billie drinks, dies. Claim: "The pouring is a cause of the drinking of a green drink, since the latter is dependent on the former; the drinking of a green drink is a cause of the death, since the latter is dependent on the former; but the coloring is not a cause of the death."

The solution is as before. Only essentially green drinkings are dependent on the coloring, while the death is not dependent on an essentially green drinking (plausibly any such drinking might fail to occur in virtue of being a drinking of another color).[18] Lewis 1973 is safe. Lewis 2000, however, is not, since undeniably the coloring influences the drinking and the drinking influences the death.[19]

To avoid these troubles, Lewis 2000 needs either modification or limitation to atomic events. I have already commented on the weaknesses of limiting analyses to atomic events (Section 4.7).

## 4.10 Michael McDermott's Analysis[20]

Michael McDermott (2002) first develops an analysis for what we may call *strict* events, events that are inflexible in every respect; he then extends it to cover our talk of flexible events. I have two serious objections to the limited analysis. I will not go

---

[18] Jig-chuen examines how various views of events might avoid his attack. But he does not consider Lewis's (1986b).

[19] (These transitive worries have led some to want to reduce causation to a relation between simple tropes, or property instances, only—for example, Douglas Ehring (1997) and Daniel M. Hausman (1998).)

[20] I discuss only the second of his two analyses. The first (1995b) is very similar but relatively complex and clunky. It clearly suffers from the same objections I will point up here. (Plus one more: an inability to handle composite events.)

beyond that analysis here, since what lies beyond does not escape those objections, and McDermott's whole approach is substantially different from the one I pursue in Chapter 5 and to the others discussed in this chapter.

First a background point. McDermott regularly talks of nonoccurrences as events, naming them, for example, "not-*c*" (so, too, in his 1995b). No definition is provided. He expresses doubt "that there is a serious distinction between events and omissions" (2002, p. 96). He takes these negative events to be causal relata relevant to his analysis. It is, for instance, only by way of them that he can count meta-preventers as causes (which he wants to do—p. 94, and 1995b): this is because his analysis relies on chains of direct causation, and there is no such chain of *positive* events connecting meta-preventer and effect (see Section 3.3.5). We could leave our understanding of these negative events at an intuitive level, or we could interpret them in some specific, appropriate way. A suggestion of Schaffer's (noted above) could serve: every event is a pair <Property, Region>, where the property may be negative—e.g., <lack of firing, neuron region>. To suppose this event to occur is to suppose there is no firing in the neuron region; to suppose it *not* to occur is to suppose there *is*.

McDermott defines a *sufficient condition* for *e* as "a condition on what happens at a point, or a conjunction of such conditions, such that given its satisfaction *e* would have occurred whatever had happened at *other* points," including past and future ones (pp. 96-7). There is no discussion of "condition" or "point." As for "point," it is clear that McDermott does not mean spacetime point—nor could he, since typically only larger direct causes are sufficient. A *minimal sufficient condition (MSC)* for *e* is "a sufficient condition in which no conjunct could be replaced by a weaker condition on what happens

at that point without losing sufficiency" (p. 97). (No discussion of, or motivation for, this definition of minimality is provided.[21]) Where *c* and *e* are distinct actual events, *c* is a *direct cause* of *e* iff *c*'s occurrence satisfies a conjunct in some satisfied MSC for *e*. (McDermott, like most writers, says "actual event" where I say "occurrent event.")

One might now take the ancestral, so that *c* is a cause of *e* iff there is a chain of direct causation from *c* to *e*. But McDermott thinks this would be a mistake. Suppose that, while I desire to press a certain button, a dog bites off my right finger. Without the bite, I would have pressed the button with that finger, but instead I use my left. The button-pressing then causes a bomb to explode (pp. 95, 99). McDermott does not want to count the bite as a cause of the explosion; but there is a chain of direct causation, as defined, connecting them: bite, pressing with left hand, explosion (McDermott is idealizing here, of course, in counting these causal connections as direct). (It may not be obvious how the pressing with the left hand satisfies a conjunct in an MSC for the explosion. McDermott does not say, but one natural suggestion is that it satisfies "the button is pressed with moderate force.") So McDermott tries a different approach.

He gives us the notion of a *sustaining condition* (p. 98). It is not precisely defined, but my understanding of it is this: suppose *c*'s occurrence satisfies a conjunct in some satisfied MSC; the remaining conjuncts in that MSC comprise a sustaining condition for *c*. If a train's going down a track is a cause of an arrival, then a sustaining condition for that cause, according to McDermott, is the presence of the track on which the train rides. In McDermott's examples, the events that satisfy sustaining conditions are all, I would

---

[21] Standardly, a (contingent) proposition P is a weakening of another Q iff Q entails P but P does not entail Q. "A concert occurred" is weaker than "A loud concert occurred." And "No concert occurred" is stronger than "No loud concert occurred."

say, joint causes with *c*. McDermott then talks of a *channel* as a series of sustaining conditions for a given chain of direct causes: channels support chains. *c* is a *cause* of *e* iff (i) there is a chain of direct causation from *c* to *e*, and (ii) had *c* not occurred, then either *e* might not have occurred at all or, if it would have, it would not have occurred by the same channel that actually supported that chain.

The sustaining condition for the bite as direct cause of the pressing was that I desired to press the button. The sustaining condition for the pressing as direct cause of the explosion was that the button, wires, bomb, were set up a certain way (p. 99). The desire and button-bomb connection together make a channel from the bite to the explosion. Had the bite not occurred, the explosion would have occurred via this same channel, so clause (ii) is not satisfied: that the bite did not occur and the desire *did* is an MSC for the pressing with the right finger, whose first conjunct is satisfied by not-bite; then that the button is pressed with moderate force and the button-bomb connections are in place is an MSC for the explosion, whose first conjunct is satisfied by the pressing with the right finger; so via the same desire–button-bomb channel, there is a chain of direct causation from *not*-bite to explosion.

One can see that, in Figure 3.1, the analysis will count *a* and *a-b* as causes of *e*. Due to the use of negative events, meta-prevention preemption (Figure 4.2) is not a problem (Section 4.5.1). (Aftereffect preemption does not arise for an analysis limited to strict events.)

Before McDermott gives his analysis, he says (p. 92) he wants to count (standard non-preemptive) overdeterminers as causes. But he never returns to show us how his analysis delivers that result, and it seems to me it does not. He imagined a case in which a pair of

(stimulatory) neuron firings (*a*, *b*) overdetermine the firing of a neuron E. One MSC for the firing of E is that a stimulatory neuron in a certain location and orientation fires, E is present, and no inhibitory events occur nearby. *a* satisfies the first conjunct, so it fits the definition of a *direct* cause. But if *a* had not occurred, E's firing would have occurred via the same sustaining conditions, namely those described by the other conjuncts in that MSC—so *a* does not come out a cause. If there is an MSC by which it does, I do not see it. (All this applies to *b*, too, of course.) Perhaps McDermott should add a disjunct to his analysis to make direct causation sufficient for causation.

Now to my first serious objection. First consider Figure 4.4: neuron A is an inhibitory neuron that fires upon C, although there is no potential firing of C to inhibit, since B does not fire.

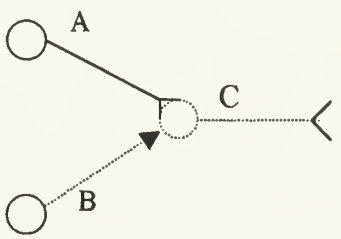


Figure 4.4

Let *a* be the (possible) firing of A, *b* of B, *c* of C. Pretheoretically, does it seem *a* is a cause of not-*c*? That an inhibitory neuron fires in such-and-such location and orientation is an MSC for not-*c* (no sustaining conditions required), and it is satisfied by *a*. So *a* comes out a direct cause of not-*c*. Incidentally, and of importance to us later, *a* also fits the definition of cause: If *a* had not occurred, not-*c* would have occurred via (different) sustaining conditions: not-*b* would satisfy (only) the first conjunct in the MSC that *b*

does not occur and no other stimulations occur. Likewise, we get that my holding up my hand (*a*) is a (direct) cause of the absence of a certain ball flight beyond my hand (not-*c*), even if no ball was thrown toward my hand (not-*b*). This is odd. We would certainly never say that just by holding up my hand I prevented a ball flight, if there were no ball around at all. It gets worse. Suppose not-*c* is a direct cause of some event *d*—perhaps an actual neuron-firing that *c* would have inhibited, or the presence of intact glass that a certain ball flight would have prevented.[22] Then *a* will come out a cause of *d*, for we just saw that without *a*, not-*c* would occur via different sustaining conditions, in which case *d* would as well (the channel to *d* would differ at *a*'s point). Likewise, my holding up my hand counts as a cause of the intact window even if no ball was thrown. Now we surely have an embarrassment of causes. Every time you walk by a window you cause its intact presence. And the presence of the book on the shelf is a cause of your breathing, by way of causing the absence of a poisoned dart between it and your heart a few moments ago—though there was no dart on the way.

My second serious complaint is that apparently the analysis is not consistent with a Humean view of laws, and it seems bad for a causal analysis to rule out such a general, and not especially implausible, view. On a Humean view of laws, the actual pattern of events in the world determines the laws: the laws hold in virtue of that pattern, such that, while laws are contingent, if worlds $w_1$ and $w_2$ have the same pattern, they have the same laws.[23] On McDermott's analysis, for an event *e* to be a direct effect, there must be an MSC, C, such that, were C to occur, *e* would occur *no matter what else happened*. But

---

[22] Not-*c* would satisfy a conjunct in this MSC for *d*: *c* does not occur and no other inhibitory event occurs.

[23] E.g.: Nelson Goodman (1955), Carl Hempel (1965), Lewis (1973a and 1994).

perhaps by way of things happening very differently throughout the world (including very relevant parts such as those properly containing conditions that duplicate C), the laws would be different enough that *e* would *not* occur, on those new laws. Perhaps there is always some MSC large enough to avoid this problem, because it covers so much of the world. Since any event whose occurrence satisfies a conjunct in an MSC counts as a direct cause, such a large MSC will have to contain conjuncts that cover large portions of the world that *include* the sorts of events we ordinarily count as direct causes—events with portions just prior to, and spatially contiguous with, *e*—if there is to be any plausibility to the claim that its satisfiers are direct causes; there will need to be MSCs that contain not only *that* sort of conjunct but contain *also* conjuncts that *merely* cover those ordinary direct causes, since we certainly want to be able to count the ordinary direct causes and therefore need conjuncts that they can satisfy. Is all this doable? It appears not. The events that satisfy the big, monstrous conjuncts covering large regions of history are in turn going to have (ordinary) direct causes far and wide, including, in many cases, in *e*'s present and future. So we can get a chain of direct causation from an event *f* in *e*'s future to *e*. And surely it will sometimes be the case that clause (ii) of the analysis is satisfied: had *f* not occurred, *e* would have occurred *not* by the same channel, in virtue of the fact that the supporting conditions by which *f* is actually a cause of the monster event would not, in *f*'s absence, be supporting conditions for any cause of it.

At this point I turn to analyses intended to accommodate both deterministic and indeterministic ("chancy") causation. Now we will be evaluating analyses not only for

their ability to handle deterministic preemption, but also for how they deal with probabilistic (potential) preemption and the problem of failed potential causes.

## 4.11 Lewis's Probabilistic Analysis

Lewis (1986d) proposed the first counterfactual analysis meant to accommodate both deterministic and chancy causation. It identifies causation with the ancestral of what the probabilistic simple analysis (PSA) identifies it with, except that in the definition of probabilistic dependence the time of counterfactual chance comparison is just after *c* instead of at the end of *c*—an unimportant difference, for us (see Section 3.2).

Let us focus on preemption first. In Section 3.4.2.1, we considered a chancy version of the situation depicted in Figure 3.1: suppose that the backup process is very reliable, while the successful process is very unreliable, save for the inhibitory power of *a*, which is at least moderately reliable. Then *a* makes *e* less likely instead of more (at the time at the end of *a*). But the ancestral analysis succeeds in counting *a* as a cause, because *b*'s chance would be lower without *a*, and *e*'s would be lower without *b* (at the appropriate times). Obviously this works also in the deterministic case, where chances are extreme. So far, so good.

However, the deterministic process preemption problem remains unsolved: If *a-b* had not occurred, the chance of *e* at the end of *a-b* might have been the same as it actually was (=1), since *f* might have occurred. And there are no events causally between *a-b* and *e* whereby the ancestral will help. And it is not just the extreme case that gives trouble—the process preemption problem carries over to the chancy environment. Suppose that each process, at any stage, is exactly as reliable as the other process at that stage. Then,

again, if *a-b* had not occurred, the chance of *e* at the end of *a-b* might have been the same as it actually was, since *f* might have occurred and *b* not.

Further, the analysis is no help with deterministic aftereffect preemption. Take any cause in Figure 4.1: where chances are all 1, if the cause had not occurred, the chance of the effect would still have been 1. Presumably Lewis intended that quasi-dependence (Section 4.5) would come to the rescue here (with probabilistic dependence in place of counterfactual dependence in its definition). But the same objections lodged above would still apply.

(*Chancy* aftereffect preemption is not an instance of the preemption problem. Where two rocks are thrown at a bottle, one getting there first, (each stage of) the first rock's journey *does* make the shattering more probable and thus does not go uncounted. Rather, the problem here is failed potential causes: early stages of the second-rock flight are also probability-raisers.)

So the ancestral of probabilistic dependence is not much help with the preemption problem. It is no help at all with the problem of failed potential causes. In a chancy setting, if *d* (Figure 3.1) had not occurred, the chance of *e* at the end of *d* would have been lower. And clearly the ancestral maneuver does nothing to help combat the overlapping variant of failed potential causes (see Section 3.4.1). (The analysis was not intended to solve the problem of failed potential causes; the problem was first raised, by Peter Menzies (1989a), in *response* to the analysis.)

## 4.12 Peter Menzies' Probabilistic Analysis

Peter Menzies 1989a is the first attempt to solve both our central problems. Causation is still the ancestral of a certain relation R. But R is now connection by an "unbroken

causal process" (p. 656): There is an *unbroken causal process* from $c$ to $e$ iff, for *any* finite sequence of n (≥0) times $<t_1, t_2, \ldots, t_n>$ between $c$'s time and $e$'s time, there is a sequence of actual events $<x_1, x_2, \ldots x_n>$ at these times (respectively) such that there is a chain of probabilistic dependence $<c, x_1, x_2, \ldots, x_n, e>$.[24] One event causes another iff there is a chain of unbroken causal processes from the one to the other.

Consider Figure 3.1 again, where the backup process is very reliable and the successful process is very unreliable, save for the inhibitory power of $a$, which is at least moderately reliable. Then $a$ makes $e$ less likely instead of more. There is no "unbroken causal process" from $a$ to $e$, because for the finite sequence of n=0 times there is not the needed dependence chain, $<a, e>$. But there is a *chain* of unbroken causal processes: where n=0, there is an unbroken causal process from $a$ to $b$, since $<a, b>$ is a dependence chain, and likewise for $<b, e>$; for these two processes, there are no times between the first and second event, so 0 is the only n value to consider. Recall that $e$ probabilistically depends on $d$, so that PSA and its ancestral variant count it a cause (Section 3.4.2.1). There is not an unbroken causal process connecting them, since at $f$'s time there is no actual event dependent on $d$ and on which $e$ is in turn dependent; and further, there is no *chain* of unbroken causal processes, as clearly no other event on the scene will serve as an intermediate link. Thus $d$ is not counted a cause of $e$. Some success, then, on both our central problems.

But the analysis is no help with process preemption. There are no events between $a$-$b$ and $e$, and $e$ is not (probabilistically) dependent on $a$-$b$.

---

[24] (Not really for *any* finite sequence of times between $c$ and $e$, surely, but rather any where $t_{i+1}$ is later than $t_i$. Otherwise there will be no causation, since there will not be a dependence chain where an $x_{i+1}$ is earlier than an $x_i$.)

The analysis falls apart, with respect to both problems, in the face of aftereffect preemption. Consider Figure 4.1. In the deterministic case, there is no dependence of *e* on *b*, and there are no intermediates (as Menzies himself (1996) points out—he there rejects this analysis[25]). Thus a preempting cause goes uncounted. In the chancy case, the failed potential cause *c* comes out a cause. *d* and *e* are simultaneous, so that *c* is just before *e*. Thus <*c*, *e*> is a dependence chain, so there is an "unbroken causal process" from *c* to *e* (noted in Noordhof 1999, pp. 99–100). (And failed potential causes of overlapping effects come out causes, as they are direct chance-raisers; see Section 3.4.1.)

## 4.13 Murali Ramachandran's Analysis

Murali Ramachandran (1997; 1998) offers an analysis that, while intended to accommodate chancy causation, makes no use of probabilities. The analysis employs the concept of a *dependence set* (*D-set*) for an event *e*, which is a set of possible events such that, had none of them occurred, *e* would not have occurred. A *minimal dependence set* (*M-set*) for *e*, is a D-set for *e* with no proper subset that is a D-set for *e*. A *temporal D-set* for *e* is a set of possible events such that, were none of its members to occur, *e* would not occur at the time it actually occurs. A *temporal M-set* for *e* is a temporal D-set for *e* with no proper subset that is a temporal D-set for *e*. The analysis goes as follows.

> For any actual events *c* and *e*, *c* causes *e* iff
> (i) *c* belongs to a temporal M-set for *e*, and
> (ii) *c* belongs to an M-set for *e*, *M*, such that for any M-set for *e*, *N*, that differs only in that it contains one or more events in place of *c*, at least one of the events replacing *c* is actual and belongs to a temporal M-set for *e*. (1998, p. 466)

[25] He says (1996, p. 96) that he had believed events have their times essentially, so that without *b*, *e* would have no chance. But he now considers this defense "ad hoc" (for reasons I do not quite understand).

Clause (i) is supposed to capture the idea that causes are relevant to the time an event occurs. Clause (ii) is supposed to capitalize on the idea, which I have been advocating, that preempted processes "do not run their full course" (1998, p. 273). The way it does this will emerge from some of the examples below.

Let us first examine how the analysis handles deterministic causation. Consider Figure 3.1. *a* is a cause of *e*. Let $M = \{a, d\}$. *M* is an M-set for *e* and also a temporal M-set for *e*. So (i) is satisfied. And *M* appears to meet condition (ii) with respect to *a*: to get an M-set, *N*, for *e* by replacing *a* in *M*, we need to include exactly one actual event from the successful process; no proper subset of *N* will be a temporal M-set for *e*, so *N* itself will be a temporal M-set for *e*. *d*, on the other hand, does not come out a cause, as condition (ii) is not met: any M-set for *e* that contains *d* will be one for which *f*, a non-actual event, can take the place of *d* to make another M-set for *e*.

The analysis is also able to count the process *a-b* as a cause of *e*: let $M = \{a\text{-}b, d\}$, and the reasoning is the same as that above for *a*.

Consider the aftereffect preemption diagram Figure 4.1. *a* is a cause of *e*. *a* belongs to a temporal M-set for *e*, namely $\{a\}$, so (i) is satisfied. Let $M = \{a, d\}$. *M* is an M-set for *e*. Any substitute for *a* will be an actual event in the successful process, and the unit set of this substitute will be a temporal M-set for *e*; so (ii) is satisfied. Precisely the same reasoning applies to the process *a-b*. *d* will not come out a cause of *e*, for the simple reason this time that it is not in a temporal M-set for *e*. The reasoning here applies exactly to the aftereffect case where two rocks are thrown at a bottle, one getting there first and shattering it.

Thus the analysis handles deterministic preemption quite well. Like PCA*, it also handles preemption by delayed action. Consider the case where *b* is removed from Figure 3.1 and *a* causes *e* at a distance. The above reasoning by which *a* is a cause of *e* when *b* is *not* removed still applies to this delayed-action version.

Ramachandran offers a simpler analysis as a stepping stone, but rejects it: (SIMPLE) *c* causes *e* iff *c* is in an M-set for *e*, *M*, such that no other M-set for *e* differs from *M* only in that it contains non-actual events in place of *c* (1998, p. 465). Consider Figure 3.1. Let *M* be {*a*, *d*}. *a* is counted a cause of *e*, because *M* is an M-set for *e* and *a* cannot be replaced by non-actual events while keeping *M* an M-set; *d*, however, can be replaced by *f*. But Ramachandran rejects SIMPLE because, for aftereffect cases like the two-rock case, where one rock arrives first and shatters the bottle, the analysis appears to count (a stage of) the second rock's journey as a cause of the shattering: let *M* be $\{s_1, s_2\}$, where $s_1$ is a stage of the first rock's journey and $s_2$ is a stage of the second rock's journey; *M* is an M-set for the shattering; in place of $s_2$, to get another M-set we can only substitute other *actual* stages involving that same rock, because no part of that journey is non-actual. Or so his reasoning goes; it is why he brings temporal M-sets to the rescue. But why not substitute the essentially late shattering that would have occurred if the first rock had not been thrown? Call this event "*h*." If *h* and $s_1$ were not to occur, there would be no shattering, and since *h* is a non-actual event, $s_2$ fails to count as a cause. We still get $s_1$ as a cause, using *M*: substitute *h* for $s_1$; it is false that without *h* and $s_2$, the shattering would not have occurred. So it seems that even SIMPLE does a good job of dealing with deterministic preemption.

In case there is some reason not to allow non-actual versions of the effect into the M-sets, let me point out that temporal M-sets are not adequate. What is needed instead are *spatiotemporal* M-sets. Let *the lunchtime ring* be an event that can occur as the ringing of either of two bells, or of both, in a certain bell tower at noon.[26] You order Lefty to ring the left bell; I give an order for Righty to ring the right bell and Molly to mute the left bell (all at noon). Consequently, Lefty's attempt is muted as the right bell rings. My order preempts yours as a cause of the lunchtime ring. Figure 4.5 may be helpful here. *r* is Righty's bell-striking, *m* is Molly's muting, and *l* is Lefty's bell-striking:

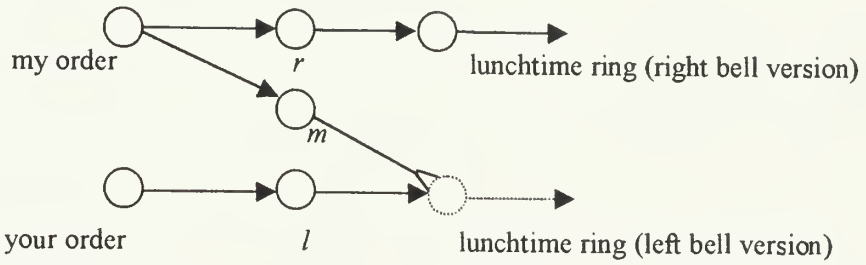


Figure 4.5

DSA fails here, because the ring is not dependent on my order. But the problem this example poses for Ramachandran, if non-actual versions of the effect are not allowed in M-sets, is that *l* comes out a cause. Let $M = \{l, \text{my order}\}$. *M* is a temporal M-set for the effect at issue. *M* is also an M-set for it. Substitutions for *l* will have to be other *actual* events in the backup process, and the result will also be a temporal M-set. So both conditions (i) and (ii) are met.

---

[26] Events have not only *temporally* flexible occurrence conditions, but spatially: the trumpet solo could have occurred as a sitting or a standing event. And occurrence conditions can be flexible with respect to object involvement: the solo could have occurred with the silver trumpet instead of the brass one. There is no motivation for counting time as the only respect in which occurrence conditions can be flexible.

The repair: change the definition of temporal M-set to read "in the region" instead of "at the time," and call it a *spatiotemporal M-set*. Just as aftereffect-preempted events are not in temporal M-sets for the effect (reflecting the fact that they do not actually influence the effect's time), so *l* is not in a spatiotemporal M-set for the lunchtime ring (reflecting the fact that it does not actually influence the effect's region).

As should be clear, Ramachandran's analysis (and each variant I have discussed) counts standard non-preemptive overdeterminers as causes. Some may take this as a good result. And one can see that it readily counts trumped events and $s_2$ in Rumpelstiltskin II as causes—some may take this as a bad result, though I have argued that it is not clearly wrong (Section 3.4.2.2).

Like PCA*, the analysis counts the backup disrupter as a cause in cases of allegedly preemptive prevention. We had the example from McDermott where I catch a ball that is headed toward a window, and between me and the window is a brick wall. I do not want to count the presence of the wall as a cause of the presence of intact glass. But my lunge and the presence of the wall make an M-set that allows both to be causes, on any of the versions of the analysis above (consider SIMPLE, for instance, and notice that there are no non-actual events that can be substituted for the wall to keep the set an M-set—flights of the ball will not do it, since in their absence, the glass is still intact). This is the main disappointment, for me, in its handling of deterministic causation.

How does the M-set approach handle chancy causation? On any of the versions we have considered, a cause must be in an M-set for the effect; but can this requirement always be met? In some cases of chancy causation, there will still be counterfactual dependence: without the presence of the radioactive nucleus, the particle emission would

not have occurred—even though it was never determined to occur. So here, the cause is in an M-set for the effect. But we can imagine two events, *a* and *b*, jointly and directly causing an event *e* by way of contributing something independently to *e*'s chance. Here, if *a* (*b*) had not occurred, *e* still might have. Are *a* and *b* in M-sets for *e*? *a* and *b* may themselves comprise an M-set for *e*, so that each comes out a cause. Ramachandran's idea is that we can always find enough events, and a minimal set of them, to be able to say that if *they* had not occurred, *e would not* have occurred.

But this is not right, for two reasons. First, it is conceivable, however unrealistic, that an event could have some background chance of occurring entirely spontaneously, with no events as direct causes, while yet it does not actually occur that way but occurs with direct event causes that make it more probable. In such a case, the direct causes will not be in M-sets for the effect. (This situation arises constantly in purely indeterministic worlds, since every effect would have some chance even if all its direct causes were not to occur.)

Second, take a standard example which is often given to motivate the need for a probabilistic analysis: a radioactive nucleus with a certain chance of emitting an alpha particle in the next minute absorbs another particle so that the chance of the emission increases; the emission occurs; the absorption seems like a cause, yet without it, the emission might still have occurred. The absorption is not enough to make an M-set for the effect. Yet anything else that is added to do it—the presence of the radioactive nucleus, for example—will by itself make an M-set for the effect, so that the absorption

will be extra and the set will not be minimal (Noordhof 1998, p. 459[27]). Ramachandran (1998, p. 469–70) holds out hope he can escape this problem if there are some events inside the atom to combine with the absorption to make an M-set. But this is a red herring, in my opinion, since conceptually the damaging situation is clear, whether or not there are real cases of it. The structure of it is simply that a cause *c* can only be in a D-set with necessary causes; the necessary causes will always make an M-set on their own, so that *c* can never get in. It can be made vivid with a simple neuron example: neurons A and B fire directly upon E; A and B are of different types, such that each one's firing makes E's firing more probable, yet A's firing is necessary in that, without it, the chance of E's firing would be 0. And intuitively we see this sort of situation day-to-day. You don't know if the small fire in the fireplace is going to grow big; to improve the odds, you add a log; the addition in fact causes a big fire, as you had hoped. The addition is in a D-set with the small fire, but it is not in an M-set: the small fire is necessary, so it makes an M-set by itself, while the log-addition does not make an M-set by itself.[28]

Consequently, I do not take the M-set analysis seriously as a broad analysis suitable to indeterministic causation.

## 4.14 Paul Noordhof's Probabilistic Analysis

Paul Noordhof's (1999) analysis is exceptionally complex. To run through the motivations behind its various components would take (as it takes Noordhof) a great deal

---

[27] Noordhof does not quite say this. He says there is no D-set for the emission that includes the absorption, since there is no D-set for the emission at all! But there are many, such as the presence of the atom.

[28] (This problem does not arise in purely indeterministic worlds, though the previous one does.)

of space. Here is the compressed motivation, supplied by the author after he presents the analysis in full (the clauses mentioned are described below):

> Although this formulation might appear complex, the final conception of a cause it articulates is relatively simple. A cause, $e_1$, is something which (independently of its competitors) both makes the chance of an effect, $e_2$, very much greater than its maximal background chance (clause (I)) *and* actually influences the probability of the effect in this way at the time at which the effect occurred (clauses (III) and (IV)) via a complete causal chain (clause (II) and the way in which probabilities are assessed). (p. 120)

"The way in which probabilities are assessed" is a reference to the fact that the chance time of counterfactual comparison is the time just before the effect. Recall, again, that in Figure 3.1, if all is chancy, the chance of *e* would have been lower at (the end of) *d*'s time had *d* not occurred, and thus PSA counts *d* a cause of *e*. But it is not true that the chance at the time *just before e* would have been lower if *d* had not occurred—it might have been the same. Thus, utilizing this chance time solves one case of the problem of failed potential causes. It does not solve them all: in Figure 4.1, *c* is a failed potential cause that makes the chance of *e* more probable at the time (at the end of) of *c*—but *this* time just *is* the time just before *e*, so assessing probabilities in the new way is no help. Noordhof is aware of this, but since the new way is partially helpful, he embraces it.[29]

The presentation of the analysis is unclear, because different, very important sentences are labeled with the same numbers in the text, and a key definition seems to be missing a label.[30] But I think the intended analysis goes as follows.

---

[29] (In Section 3.2, I describe a problem with using the time just before the effect. Whether this problem arises in the context of Noordhof's analysis I shall not investigate.)

[30] Clause (I), below, appears to be given on pp. 109-10, though "(I)" labels other sentences in the paper. The label Noordhof actually uses for the second clause below—"(II)'"—is used for a different definition as well, on p. 114.

For any actual, distinct events $e_1$ and $e_2$, $e_1$ causes $e_2$ (if and)[31] only if there is a (possibly empty) set of possible events $\Sigma$ such that (I)-(IV) is true:

(I) $e_2$ is probabilistically $\Sigma$-dependent on $e_1$:
  (1) If $e_1$ were to occur without any of the events in $\Sigma$, then for some time $t$, it would be the case that, just before $t$, p($e_2$ at $t$) $\geq x$;[32]
  (2) if neither $e_1$ nor any of the events in $\Sigma$ were to occur, then for *any* time $t$, it would be the case that, just before $t$, p($e_2$ at $t$) $\leq y$;
  (3) $x$ is much greater than $y$.

(II) For any superset $\Sigma^*$ of $\Sigma$, if $e_2$ probabilistically $\Sigma^*$-depends upon $e_1$, then every event upon which $e_2$ probabilistically $\Sigma^*$-depends is an actual event.

(III) $e_2$ occurs at one of the times for which p($e_2$ at $t$) $\geq x$ and $x$ is much greater than $y$.

(IV) $e_2$ probabilistically $A$-time depends upon $e_1$:
Where $t_0$ is $e_2$'s actual time, there is a possibly empty set of possible events $A$, meeting constraint (C), below, such that
  (1) If $e_1$ were to occur without any of the events in $A$, then it would be the case that p($e_2$ at $t_0$) $\geq x$;
  (2) if neither $e_1$ nor any of the events in $A$ were to occur, then it would be the case that p($e_2$ at $t_0$) $\leq y$;
  (3) $x$ is much greater than $y$.

(C) No event $e_i$ is in $A$ if both
  (a) If it is a member of $A$, then $\langle e_1, e_2 \rangle$ satisfies (IV),[33] and
  (b) if it is not a member of $A$, and we replace (IV)(1) and (2) with
    (1*) If $e_1$ and $e_i$ were to occur, with none of the events in $A$ occurring, nor $e_i$ satisfying any of (I)-(III) regarding $e_2$, then it would be the case that p($e_2$ at $t_0$) $\geq x$
    (2*) If $e_i$ were to occur with neither $e_1$ nor any of the events in $A$ occurring, nor $e_i$ satisfying any of (I)-(III) regarding $e_2$, then it would be the case that p($e_2$ at $t_0$) $\leq y$,
  then $\langle e_1, e_2 \rangle$ does not satisfy (IV). (pp. 109–10, 116–7, 120)

---

[31] Noordhof uses parentheses here because he is not confident that the analysis is—or that extant counterfactual analyses generally are—generally asymmetric (pp. 120ff). He doubts Lewis's claim that backtracking counterfactuals are special, are standardly false (see Section 2.2).

[32] "p($e_2$ at $t$)" means the probability of the proposition *that $e_2$ occurs at $t$*.

[33] (IV) save "meeting constraint (C), below," we must assume, else we are caught in a circle here.

I think the analysis is *unacceptably* complex. The clauses have a certain common style, but they are not common enough to defeat the charge of complexity—the analysis is too long, fragmented, and difficult. Despite its motivations, it is not intuitive enough for me to believe that people (and children) have anything like it in mind, however tacitly, when they employ the notion of causation. It is implausible that our sense that causation is a fairly simple concept is *this* far off. And it is, I dare say, impossible to convince oneself that the probability of serious counterexamples is low, especially once one has seen how counterexamples to much simpler analyses can lay hidden.

I shall not attempt to evaluate how well the analysis handles our two central problems, because I do not think it is worth it. Even if it handles them (or *appears* to handle them), a much simpler analysis will be wanted.[34]

## 4.15 Conclusion

For deterministic causation, PCA* and the M-set analysis get good results overall. But they count the second potential disrupter in allegedly preemptive prevention cases as causes, which I think is a bad result. Also, I believe I can produce something more elegant and intuitive than these analyses.

As for indeterministic causation, it appears that no counterfactual analysis to date has come close to being acceptable. It may be necessary to proceed more slowly, to retreat to a special case and see if we can produce a decent analysis at least for *it*; this is what I am aiming for in Chapter 6.

---

[34] The analysis does appear to suffer from the overlapping problem. Noordhof admits (p. 100–8) that he is relying on the idea that all cases of failed potential causes are cases involving failed intermediates; but overlapping cases do not. Intermediates and time issues are irrelevant to overlapping, and those are the only issues his analysis addresses.

# CHAPTER 5

# A NEW DETERMINISTIC ANALYSIS (AND TWO SIMILAR RIVALS)

## 5.1 Introduction

In this chapter, I propose my own analysis of deterministic event causation. Recently two similar analyses have been published (Christopher Hitchcock 2001 and Stephen Yablo 2002). In commenting on these views I want to draw comparisons to and contrasts with my own, so I have reserved discussion of them for after I present mine; this will come at the end of the chapter.

## 5.2 The Analysis

Consider again the preemption situation depicted in Figure 3.1, and assume determinism.

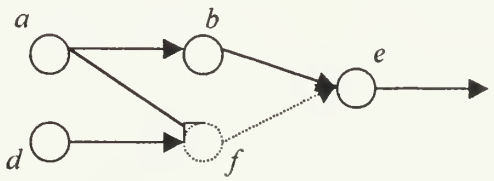


Figure 3.1

*a* is a cause of *e*, but *e* is not dependent on *a*. However, notice that if *a* were not to occur and (still) *f* were not to occur, *e* would not occur. Similarly, if *a*-*b* were not to occur and (still) *f* were not to occur, *e* would not occur. For the analogous dart–balloon story (Section 4.2), where my dart-throw (*a*) preempts Lucy's windup (*d*) as a cause of the pop (*e*), we can say likewise: Were my throw not to occur and still Lucy's throw (*f*) were not to occur, then the pop would not occur. And if the whole journey of my dart (*a*-*b*) were

not to occur and still Lucy's throw were not to occur, then the pop would not occur. By holding fixed that a certain event does not occur (a certain part of the backup process) we seem to regain a kind of dependence of effect on cause.

This holds for the case where *a* causes *e* by delayed action, and there is no intermediate, *b*: if *a* and *f* were not to occur, *e* would not occur.

Consider the lunch-bell example (Figure 4.4) in which I issue an order for Righty to ring the right bell and Molly to mute the left bell, thereby preempting your order to Lefty to ring the left bell. If my order had not occurred and (still) the left bell had not rung, there would have been no lunchtime ringing. Likewise, if the process from my order up through Righty's bell-striking had not occurred, yet the left bell were not to ring, there would be no lunchtime ringing.

Consider the dinner-bell example in which I give an order for Amy to ring the bell at 5:59 and Betty to remove it at 6:00, thereby preempting your order to Carl to ring it at 6:01. If my order were not to occur, and there were no 6:01 ringing, there would be no dinnertime ringing. This is as true of the process from my order up through Amy's bell-striking.

We can say analogous things in the aftereffect preemption cases. Where the 5:59 ringing *causes* the absence of the 6:01 (by causing Betty to remove the bell), it is no less true that without my order and without the 6:01, there would be no ringing; and, without the Amy process and without the 6:01, there would be no ringing. In the two-rock example, there is a merely potential shattering (*h*) that essentially occurs at or shortly after the time the second rock would have hit the window had the first rock not been thrown. If the first rock throw were not to occur and yet *h* were not to occur, then the

shattering would not occur—whether this "throw" is a stage of the rock's journey or the whole process.

Consider the example of meta-prevention preemption where my untying the safety-net rope is a cause of the trapeze artist's death, and the untying preempts someone else's cutting of the rope (Figure 4.2). If the untying hadn't occurred, and still the cutting had not occurred, the death would not have occurred.

We might describe this regained dependence as counterfactual dependence holding certain things fixed. The "fixing," thus far, has worked like this: instead of saying, "if *a* were not to occur, *b* would not occur," we hold fixed the nonoccurrence of events $f_1, f_2$, ... by saying, "if $a, f_1, f_2$, ... were not to occur, *b* would not occur." But let us notice another kind of fixing that needs doing. Consider the following example.

Eleven people are on a sinking ship. Crewman Jill tells Captain Jack she thinks the raft can handle only ten. Jack is skeptical, but unsure. He tells her, "If I'm not on the raft in 5 minutes, go without me." Jill promises herself that if Jack *does* get on the raft, *she'll* stay with the ship instead. But Jack doesn't. Jill and the nine passengers wait out the 5 minutes, then float to safety. Jack is swallowed by the sea. The survivors eventually learn (from the raft manufacturer) that indeed the raft would have sunk with 11 people. The passengers also learn of Jill's willingness to stay with the ship, and they thank her for that. They say a prayer for Jack and thank him in their hearts; they say they owe their lives to Jack; they wish he knew that he made a worthwhile decision, that it wasn't a waste after all, and that its end was achieved; they tell their grandkids how Jack saved their lives.

These people view Jack's decision to stay with the ship as a cause of their survival, of their safe arrival home; I agree. Now, if Jack's decision hadn't occurred, Jill would have stayed back, true to her promise; so the arrival wasn't dependent on Jack's decision. Jack's decision preempted Jill's promise as a cause of the arrival. However, if Jack's decision hadn't occurred *and still Jill had remained on the raft*, the arrival wouldn't have occurred, because the raft would have sunk. This preemption challenge is met by holding fixed an *occurrence* in the preempted process—Jill's sitting in the raft. It is by preventing oneself being on the raft that one's decision or promise has the potential to be a cause of the arrival. The Jill process is failed because Jill is *on* the raft, and this is what we need to hold fixed.[1] In general, hold fixed the occurrence of events $f_1, f_2, \ldots$ by saying, "if *a* were not to occur and $f_1, f_2, \ldots$ *were* to occur, *b* would not occur."

In these examples, while things are held fixed, not only would the effect not occur without the cause, it *would* occur *with* the cause—as it does. The cause and the fixed background constitute a situation given which the effect would occur, and the cause is necessary for that occurrence in the context of that situation: the cause is (counterfactually) necessary as part of a situation (counterfactually) sufficient for the effect. Counterfactual dependence is a limiting case: the cause itself comprises a situation given which the effect would occur, and it is necessary for the effect. Thus both preemptive and non-preemptive causes are necessary as parts of situations in which the effect would occur. Now, it may seem that the trick at this point is to define the sort of situation that can play this role—to say what can and cannot be held fixed. But I doubt

[1] A neuron diagram will reveal that no nonoccurrence can be held fixed to gain dependence of effect on cause. (I spare you that diagram, due to its complexity.) But regardless, I hope you agree that the fact that Jill is on the raft is a fairly natural choice for something to hold fixed to gain dependence.

there is any such need. My proposal, roughly, is that one event causes another iff it is necessary as part of *some* situation sufficient for the other. Later on, I will try to allay fears that this is overly permissive. Right now, some precision.

Let a *simple event proposition* be any proposition that a certain event occurs or that a certain event does not occur, to be symbolized, for example, O($e$), ~O($e$). (An *event proposition* is any result of Boolean complications of simple event propositions.) Let a *situation* be any non-empty set of simple event propositions whose involved events are all essentially distinct. A set *holds* iff all its members are true propositions; it *implies* whatever the conjunction of its members implies (∅ always holds, never implies). An *actual situation* is one that actually holds. Call a situation's members its *parts*; and say an event $e$ is *part of a situation* $S$ iff O($e$) is part of $S$.

Then, $b$ is *situationally dependent* on $a$, in virtue of situation $S$ iff $a$ is part of $S$ and (i) if $S$ were to hold, $b$ would occur, but (ii) if $S$–{O($a$)} were to hold and {O($a$)} were *not* to hold, $b$ would *not* occur. The antecedents of (i) and (ii) imply that $S$–{O($a$)} holds; this set, when nonempty and actual, is a situation that represents what is "held fixed"; the things we held fixed in the examples above belong in such a set. The limiting case arises when $a$ is necessary as part of the situation {O($a$)} sufficient for $b$; iff that is so, $b$ is counterfactually dependent on $a$. I propose that $c$ is a *cause* of $e$ iff $e$ is situationally dependent on $c$ in virtue of some actual situation. More colloquially: iff $c$ is necessary as part of some actual situation in which $e$ would occur.

For all cases where $b$ is situationally dependent on $a$ in virtue of a situation $S$ that is not identical to {O($a$)}, we could say that $b$ is thus *conditionally dependent* on $a$: $b$

depends on *a given* $S$–{O($a$)}. If instead $S$ is {O($a$)}, then we can say $b$ is *outright dependent* on $a$. Situational dependence is dependence either outright or conditional.[2]

As far as I can see, every case of outright dependence is also a case of conditional dependence: there is always *some* occurrence or nonoccurrence that can be taken as given. In fact, there are always *actual* ones that can be taken as given (held fixed): the occurrence of some event in the distant past, or the nonoccurrence of some unlawful event (if I were to flip the switch, the light would go off; but then it is also true that if I were to flip the switch and not suddenly turn to stone, the light would go off).

Therefore, not only in preemption but also in non-preemptive causation we can find some *actual* situation $F$ (distinct from cause $c$) such that $e$ depends on $c$ given $F$. Since $F$ is actual, it is natural in such cases to say, "holding certain things fixed, $e$ wouldn't have occurred without $c$," or, "$e$ is dependent on $c$ certain other things being equal." (E.g., "Certain other things being equal, the balloon wouldn't have popped if I had not thrown.") Then we can amend DSA with just this kind of qualifier: $c$ is a cause of $e$ iff they occur, they are essentially distinct, and $e$ is dependent on $c$ *certain other things being equal*. This analysis is equivalent to the one in terms of situational dependence, thanks to the fact that we can always find something actual to hold fixed. Both are simple and intuitive and may correspond to ways we think about particular cases.

Our causal intuitions differ in character according to whether causation holds in virtue of outright dependence or only conditional dependence. Consider the bell stories of the

---

[2] The definition of conditional dependence here can be taken to descend from a broader conception of it that may perhaps have some use in other arenas: given P, (i) if Q were so, R would be, and (ii) if Q were *not* so, R would *not* be—for *any* propositions P, Q, R. (That is, if P and Q were so, R would be, and if P were so but Q were not, R would not be—for any P, Q, R.)

previous chapters: we feel that my order is in some sense not really helping; it hampers as much as helps, by blocking the other process, and consequently it makes no net difference to whether the effect occurs. We could capture this by saying the order is *not* a cause *outright*; but given certain of the circumstances (the failure of the backup ringer), the preempting cause *is* a help—it *is* a cause *conditionally*. The distinction between the two kinds of dependence provides an analysis of the intuitive causal distinction just noted. The analysis explains our inclination to see *preempting* causes as a little less deserving of the name "cause": only one of the two kinds of dependence holds; and, this kind would seem to be the less basic or fundamental.

### 5.3 Comparison to John L. Mackie's Theory

The proposal is reminiscent of John L. Mackie's (1965), though for sure they are also importantly different.[3] According to Mackie, a cause is a member of every actual, minimal lawfully sufficient condition for the effect (though there are some further requirements, the most important of which pertains to asymmetry). Thus, as Mackie says, a cause is a necessary part of a sufficient condition for the effect. Mackie and I seem to mean roughly the same thing by "condition" and "situation" (he does not define "condition," but he composes conditions by conjoining propositions that roughly correspond to occurrences and nonoccurrences). Mackie's condition is *lawfully* sufficient for the effect, while my situation is counterfactually (or subjunctively) sufficient; but in fact Mackie's condition is sufficient in *both* senses, since a condition that implies *e* is also

[3] The similarity between Mackie's view and dependence certain-other-things-being-equal was pointed up to me by Phillip Bricker, which observation motivated the situational-dependence picture, necessity as part of a situation in which *e* would occur.

one given which *e* would occur, and, my analysis is equivalent to one using situations that also are sufficient in both senses (at least for worlds with laws disallowing action at a distance). How so?

If *b* is situationally dependent on *a* in virtue of an actual situation, it will be so in virtue of many actual situations, and some of these will lawfully imply that *b* occurs. For instance, for Figure 3.1, *e* is situationally dependent on *a* in virtue of the situation $S=\{O(a), \sim O(f)\}$: if *a* were not to occur and still $S-\{O(a)\}$ were to hold, *e* would not occur. If we add to *S* enough to characterize just the state of the world at (the beginning of) *a*'s time, then the result, $S^*$, will lawfully imply that *e* occurs (since the world is deterministic), and it will be true that if *a* were not to occur and $S^*-\{O(a)\}$ were to hold, *e* would not occur—it will be true because its being true for $S-\{O(a)\}$ did not hang on it being the case that certain *a*-simultaneous extras would not or might not obtain in the antecedent situation, and thus adding that they would hold is innocuous. So situational dependence between occurrent, essentially distinct events arises only if it arises with respect to a situation that lawfully implies the effect. And thus without loss the analysans can be restricted to such situations: A cause is counterfactually necessary as part of a situation that lawfully implies the effect. It may be that in thinking of an event as a cause, we often have some of this larger, lawfully sufficient situation in mind: *The balloon was still at last, I had a nice clear line of sight to it, and my dart was newly sharpened; I threw the dart, and Lucy didn't throw hers—clearly, then, my throw was a crucial ingredient in bringing off the pop.*

There is, of course, a big difference in the ways the analyses count the cause "necessary." For Mackie, it is necessary in that the condition to which it belongs is

*minimally* lawfully sufficient for the effect; for me, it is necessary in that, with its absence, in conjunction with the balance of the situation, the effect would not occur.[4]

## 5.4 Defensive Comments, Concessions

As I said, one would have expected that we needed to specify *which* occurrences and nonoccurrences can be held fixed, rather than allowing *any* (apart from *c*) to be fixed. Have I succeeded in counting lots of preempting causes by being overly permissive? I consider three ways my proposal may appear to overcount causes. Then I return to some of the undercounting allegations made against DSA, which could as easily be made against my view; we need to see whether there are explanations for (what I claim are) our erroneous causal judgements in those cases that are consistent with my analysis.

### 5.4.1 Movers as Causes[5]

A safe is falling towards a jelly doughnut on the sidewalk. I bump the doughnut left a few inches, so Squash occurs in spatial region S instead of R. Let R-squash and S-squash be squashes that essentially occur in R and S, respectively (R-squash being nonoccurrent). Squash, the effect at issue, is more flexible and could occur in either region. Now, we might say this: (Q) "If the bump hadn't occurred and (still) R-squash hadn't occurred, Squash wouldn't have occurred." If Q is true, the bump is a cause of Squash. And that result may look wrong.

---

[4] It may be that a version of my analysis restricted to situations that are minimally lawfully sufficient is also true. If so, my view is Mackie's plus the counterfactual necessity just mentioned. Whether that is so or not, we can see that the counterfactual does all the work for my analysis, since I do not *need* to add a requirement of either lawful sufficiency or minimal lawful sufficiency.

[5] Here I rely more heavily than elsewhere on your familiarity with the semantics of counterfactuals; see Section 2.2.

I do not find Q compelling. To say that, in the antecedent situation, Squash wouldn't have occurred at all, is to suggest that it would have been that, say, the safe did not fall, or did not fall near there, or the doughnut was absent or in a region outside Squash's essential range. But if those things might have been, then why might the doughnut not also have been in S? There is a slight arbitrariness in ruling that out, it seems to me. So I have a hard time accepting Q.

But suppose the doughnut is on a train-track rail, instead (let this be region "R" now); just before the train arrives, I bump the doughnut over onto the other rail (S), where it is squashed. Consider Q again, applied to this new scenario. I think it is a reasonable claim. In the absence of the bump and R-squash, it would be that either the train stopped, or the doughnut was off the rail; but its being on the *other* rail seems a little fanciful, a gratuitously neat hypothesis, a bit far-fetched.

The analysis is flexible, then, with respect to this sort of case. It counts the mere spatial mover from R to S a cause when the movement accomplishes a substantial change, a change that, without the mover, we cannot even say *might* have occurred in the absence of the R version of the event.

Probably some will be unhappy counting the mover a cause even in such cases as the rail version. I concede I am not sure it is a good result; but I think the flexibility captures a genuine variance in our causal intuitions: in general a mover seems more responsible for the (flexible) effect when it is responsible for a version that seems "a far reach" from the actual version.

Note that the analysis definitely excludes some movers as causes. Causation arises only when the relevant counterfactual-antecedent situation would leave the consequent

event without sufficient parts (parts whose occurrence entails the occurrence of the event). Let us assume that the Great Chicago Fire could have occurred a little differently, including a little smaller or larger, a bit to the north or south. Imagine that a gust of wind in fact caused the whole fire to occur a bit south of where it otherwise would have, say one hundred meters south. If the gust had not occurred, and still there were no fire in that hundred-meter northern strip, the Fire would still have occurred—because the portion below the strip that would still have occurred is a sufficient part of the Fire. So the gust will not come out a cause of the Fire. This point will apply to temporal as well as spatial movement, of course. Putting these sorts of very slight movements aside, let us now consider temporal movements.

Suppose I push the top of the safe as it falls so that it falls faster, and consequently Squash occurs earlier (in R). It seems that if the push were not to occur, and (still) the late squash were not to occur, Squash would not occur—in the antecedent situation, the safe would fall at the original, slow rate. Plausibly, Squash occurring *earlier* in the antecedent situation is more far-fetched than its not occurring at all. So the analysis seems to count hasteners as causes. This result, too, may seem wrong. I am not sure this is a good causal result, but perhaps it is fine. I sympathize somewhat with those (Section 4.5) who think hastening seems like causing: a hastener induces the event, it brings it on.

In general the situation with respect to delaying is more akin to that of spatial moving (and my opinion of it is the same): the analysis seems somewhat flexible. Suppose that a swarm of bees slows a rock's journey to a bottle just a little. If the swarm hadn't occurred and still the early shattering hadn't occurred, ...well, one *might* say that the late shattering might have occurred instead (the window might have hesitated before

shattering). If the delay is great, however, it would seem far-fetched that such a late shattering might occur. Aside from this general similarity with spatial moving, in delaying there appears to be a special sort of example where the delayer is decidedly excluded as a cause no matter what the size of the delay. The example comes from Ned Hall (2000, p. 208): Billy and Suzy are winding up to throw rocks at a bottle; Suzy is the faster throw, but a sudden cramp in her arm stops her from throwing; Billy continues on, causing the shattering. The cramp delays the shattering. But on the proposed analysis it is not a cause, for there is nothing one might hold fixed such that without the cramp the shattering wouldn't have occurred. It is definitely false, for example, that if the cramp hadn't occurred but still there were no early shattering, the shattering wouldn't have occurred. Figure 5.1 diagrams the example.

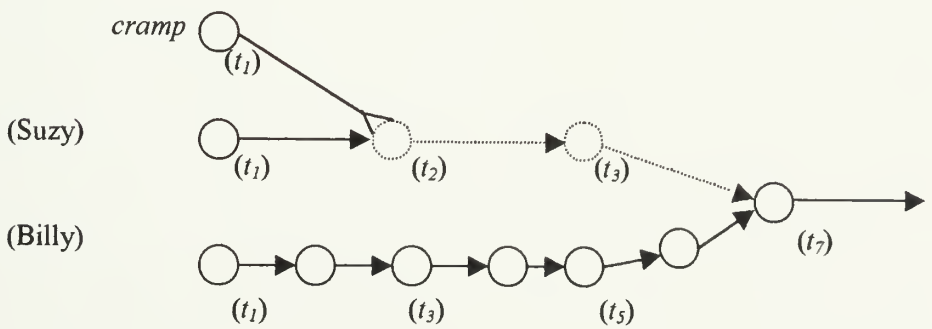


Figure 5.1

5.4.2 Threat/Savior Cases

Two red lanterns appear on a boat in the harbor, and they are visible from the castle. The royal cook sees them and thinks they are a signal from his rebel gang that he should poison the king's drink now, so he does. The king's chamber guard also sees the lanterns, but he thinks they are a signal from fellow guards that someone may be trying to

poison the king and he should put antidote in any food or drink brought to the chamber, and so *he* does. The king drinks, but lives. Is the presence of the lanterns a cause of the king's being alive? Do they save him? I prefer to say "Yes." As with preemptors, the lanterns both help and hinder. They help in that they cause the guard to add antidote to the drink; and I think it is fair to say that they help the guard to save the king's life by alerting him to the possibility of poison in the drink. Once I see this helpfulness, I prefer to count the lanterns a cause of the survival.

That sort of example is sometimes put forward as a counterexample to transitive causal analyses (such as Lewis 1973, Ganeri *et al.* 1998, Lewis 2000—see Chapter 4): the lanterns cause the addition of antidote, which in turn causes the king's live presence, but the lanterns (it is said) do not cause the king's live presence.[6] The situational-dependence analysis also counts the first a cause of the third: if the lanterns hadn't been present and yet still there were poison in the drink, then the king would have died. This particular example is one I devised to try to show that it is fine to count the first a cause of the third. But other examples perhaps make the causal claim harder to swallow. Consider the following. The cook decides on his own to poison the king. The guard sees the cook add poison to the drink, and so he adds antidote.[7] Did the addition of poison save the king's life? Was it a cause of his later live presence? The structure is basically the same here as in the example above. But perhaps it is harder here to answer Yes.

---

[6] Ned Hall (1997, 2000) and Stephen Yablo (2002) present this kind of challenge. (Hartry Field is often credited with early instances of this sort of example (see Yablo 2002, p. 134).) The examples in Hall 2000 have a special structure in which the final event is clearly not situationally dependent on the first; this is made clear by way of Christopher Hitchcock's discussion (2001, Sections 7–9).

[7] This particular example is from Hitchcock (in correspondence).

I think this may be because, in the second example, the poisoning's role as threat is more salient than its role as savior. The reason for this is probably that we do not think of an act of poisoning as primarily a reflector of light images; but in the story, this is one of the things it is—it sends light into the guard's eyes, signaling him that there is poison in the drink. In this way, it saves the king. Focus on the fact that the poisoning event is not only a poisoner of liquid but a transmitter of a warning message, and the asymmetry tends to even out.

### 5.4.3 Switching

Another sort of alleged counterexample to transitivity goes as follows. The train track splits and the two branches reconverge before the station. A train approaches the fork; the switch is thrown so that the train travels on track B instead of A; it arrives at the station just as it would have without the switching. The switching is a cause of the B journey, and the B journey is a cause of the arrival, but to many it does not seem that the switching is a cause of the arrival. The situational dependence theory yields that result as well: if the switching had not occurred and still the A journey had not occurred, then the arrival would not have occurred. (Perhaps the counterfactuals here are more clearly true if the branches spread far apart and "A journey" and "B journey" are taken to be intermediate stages of those journeys, close neither to the point of divergence nor to the point of convergence. One may add those stipulations.)

I agree that it is a bit of a strain to count the switching a cause of the arrival. But I think the verdict may be correct. The switching prevents one journey but causes another that results in the arrival. Like the preemptors we have seen, it helps and hinders, but it still helps. It is an example in which the lack of outright dependence is especially salient,

more so than the presence of conditional dependence. This would be alleviated some if we called the switching something else: route positioning, route locating, channel opening. It opens one channel and closes another. It is an opening/closing. Was the opening/closing a cause of the arrival? Yes and No. Conditionally, but not outright.

Here is a little thought experiment that may make it seem more reasonable to count the switching a cause. Imagine you know that the train is approaching the fork, that the switch is set to track A, but that the train does not travel on track A. Then you learn that the train arrives at the station. How could it? The switching provides an explanation.

### 5.4.4 Undercounting Challenges Revisited

For "Rumpelstiltskin II" and trumping (Sections 3.4.2.2.3 and 3.4.2.2.4), we allegedly had preemption without failure in the backup process, and I claimed that the alleged preemptor is not really a cause. It would be good to have an explanation as to why we commit this error in judgement, an explanation consistent with the situational-dependence analysis of causation. I said (in those earlier sections) that the alleged preemptor does not clearly look like a cause once we see that its complete circumstances are already "enough" for the effect. But we should ask, "Why, if causation is what I say it is, do we prefer to count those cases as preemption when we ignore this fact about the sufficiency of the circumstances, and why does attending to the fact put a damper on that preference?"

Here is a theory that seems plausible to me. In the examples at issue, it seems offhand that the alleged preemptor *c*'s circumstances fail to imply the effect. As a result, it feels intuitively that those circumstances must have, or lack, some intrinsic feature in virtue of which they fail to imply it. Then, it seems that holding fixed this impotent nature, the

effect would not occur without $c$ (that is, we hold fixed a version of the circumstances that essentially has this nature). But to the extent we can absorb the fact that the circumstances do imply the effect, there is no impotence to hold fixed, and no situational dependence.

For standard non-preemptive overdetermination (Section 3.3.4), I simply said that there are intuitions both for and against counting overdeterminers as causes, and that therefore an analysis could deliver either verdict acceptably. My analysis counts the overdeterminers non-causes. Now, if causation is situational dependence, and there is no situational dependence of effect on overdeterminer, then why is there a significant inclination to count the overdeterminer a cause? I do not know. Until there is an answer, this stands as something of a failing.

## 5.5 Christopher Hitchcock's Analysis

Christopher Hitchcock's analysis (2001) uses a kind of causal modeling borrowed from the work on causation of computer scientists Joseph Halpern (2000) and Judea Pearl (2000).[8] Hitchcock takes a *causal model* to be a pair of sets $V$ and $E$. $V$ is a set of variables, each of which corresponds to an event: a variable may be binary, in which case a value of 1 means that the corresponding event occurs, 0 that it does not; or a variable may be multivalued, in which case its values are particular (not-too-far-fetched) alterations of it. The notion of an alteration is taken from Lewis 2000 (Section

[8] Pearl (2000) has a counterfactual analysis of particular-event causation that, as Hitchcock notes, is similar to Hitchcock's. But it is more complex than Hitchcock's, and what I have to say about Hitchcock's would, I believe, carry over to it anyway, so I have not taken it up here.

4.9, above): an *alteration* of an event $e$ is either an event with strict occurrence conditions that is similar to how $e$ actually occurred, or it is the nonoccurrence of $e$.[9]

$E$ is a set of *structural equations* relating the values of the variables in $V$. Structural equations come in two forms. (Variables are represented by italicized capital letters.) Some take the form $X = x$—they merely state the actual value of the variable. Others express the value of a variable as a function of the values of other variables (in $V$): $Z = f_Z(X, Y, \ldots, W)$. Such an equation encodes a set of counterfactuals of the form, "Were it the case that $X = x$, $Y = y, \ldots, W = w$, then it would be the case that $X = f_Z(x, y, \ldots, w)$." (This is why the equations are "structural"—unlike mathematical equations, they are not reversible.) Thus the equation reveals how $Z$'s value would vary given variations in values of the variables on the right side of the equation. (For this sort of equation, the variables $X, Y, \ldots, W$ are called *parents* of $Z$.) The correct equation of this sort for a variable $Z$ can be constructed by first expressing the value of $Z$ as a function of all other variables in $V$ and then deleting those variables whose values make no difference to $Z$ given every assignment of values to the other variables (p. 281). The equation for $Z$ will thus tell us precisely which variables $Z$ is counterfactually sensitive to given certain values of other variables in $V$. (Every variable in $V$ is to appear on the left-hand side of exactly one equation in $E$.)

As an example, Hitchcock models a preemption story like the one above in which Lucy refrains from throwing her dart because she sees me throw—without my throw, the

[9] Lewis calls them "absences," not "nonoccurrences." For Lewis, to say that an absence (alteration) occurs is just to say that the event does not occur; so an absence, or nonoccurrence, is not itself an event. But Hitchcock does not comment on what it is (or isn't).

balloon-pop would have occurred anyway, by way of Lucy's throw. Let the variable $C$ be 0 or 1 according to whether I throw my dart; $L$ is 0 or 1 according to whether Lucy throws; $E$ is 0 or 1 according to whether the pop occurs. The set of equations can be this: $C = 1$; $L = 1–C$; $E = \max\{L, C\}$. The first says that I throw. The second says that the value of L would be 0 if C were 1, but 1 if C were 0—that is, Lucy would throw if I were not to, but she wouldn't if I were to. The third equation says that $E$'s value would be 1 if either $L$'s or $C$'s were 1, but 0 if $L$'s and $C$'s were both 0—that is, the pop would occur if either Lucy or I were to throw, though not if neither were to.

A *route* between two variables $X$ and $Z$ is a sequence of variables from $X$ to $Z$ where each member of the sequence is a parent of the next. Hitchcock sometimes refers to this as a "causal route." A variable is *intermediate* between $X$ and $Z$ iff it is distinct from $X$ and $Z$ and it belongs to a route between $X$ and $Z$. In the model just described, we can see two routes from $C$ to $E$: $<C, E>$ and $<C, L, E>$. In the second, but not the first, $L$ is intermediate between $C$ and $E$.

The notion of an *active route* is central:

> The route $<X, Y_1, \ldots, Y_n, Z>$ is active in the causal model $<V, E>$ iff $Z$ depends counterfactually upon $X$ within the new system of equations $E'$ constructed from $E$ as follows: for all $Y \in V$, if $Y$ is intermediate between $X$ and $Z$, but does not belong to the route $<X, Y_1, \ldots, Y_n, Z>$, then replace the equation for $Y$ with a new equation that sets $Y$ equal to its actual value in $E$. (If there are no intermediate variables that do not belong to this route, then $E'$ is just $E$.) (p. 286)

The idea is that a route from $X$ to $Z$ is active if, perhaps holding fixed certain actual values of variables along *other* routes from $X$ to $Z$, $Z$'s value would have differed if $X$'s had. In the model above, we can see that $<C, E>$ is active. Notice that $L$ is intermediate

between $C$ and $E$ but does not belong to $<C, E>$. So we replace the equation for $L$ with "$L = 0$." Since now $L$ equals 0, $E$ equals 0 or 1 according to whether $C$ equals 0 or 1 (because $E = \max\{C, L\}$)—and this is the "counterfactual dependence within the new system of equations." One can see that the activity of $<C, E>$ is revealed by this pair of ordinary-English counterfactuals: "If Lucy were not to throw and I were to throw, the pop would occur"; "If Lucy were not to throw and I were *not* to throw, the pop would *not* occur." These are, of course, just the pair in virtue of which the pop is situationally dependent on my throw.

Where variables $C$ and $E$ correspond to distinct events $c$ and $e$, $c$ is a *cause* of $e$ iff there is an active causal route from $C$ to $E$ in an appropriate causal model. Appropriateness has "at least" three components (p. 287): the equations in the model must entail no false counterfactuals; the dependence relations represented must be between distinct events (no definition of distinctness is given); and where the variables are multivalued, their values should not be alterations that we consider far-fetched.

In the multivalued case, it is clear that one of the objections I lodged against Lewis 2000 (Section 4.10) would also apply: the throw of the rock that slips through the hole in the glass comes out a cause of the shattering, since alterations of it would yield alterations in the shattering. However, objections that turned on Lewis's use of an ancestral relation would not. For instance, on both analyses, putting food-coloring in the poisoned drink comes out a cause of the drinking of a green poisoned drink (close alterations in the former yield alterations in the latter), and this drinking in turn comes out a cause of the death; Lewis's transitive analysis will therefore count the addition of food-coloring a cause of the death; but Hitchcock's will not, since there is nothing $F$ to hold fixed such

that if close alterations of the coloring occurred but still $F$ held, non-actual alterations of the death would have occurred.

As for the binary case, the analysis cannot clearly handle aftereffect preemption. For there are cases where there seems to be nothing to hold fixed to gain dependence other than the nonoccurrence of the late version of the effect, and—as I will explain—Hitchcock's analysis does not allow this. Let me give a precise illustration of the sort of case I mean—call it an instance of *tight aftereffect preemption*. Suppose time is discrete. Let $e$ be the firing of "neuron" N, an event that could occur at $t_2$ or $t_3$. Say that, by law, upon being stimulated neurons fire for one unit of time, whereupon they vanish (at the next time they do not exist). Imagine that at $t_1$, N is directly fired upon ($c$) by one (stimulatory) neuron, and N fires ($e$) in reaction immediately after, at $t_2$. However, also at $t_2$, yet another neuron fires directly upon N; but N does not fire at $t_3$, since N is gone then. So $c$ is a cause of $e$, but without $c$, $e$ would have occurred anyway, at $t_3$. If $c$ had not occurred and yet N's firing at $t_3$ had still not occurred, $e$ would not have occurred—by holding fixed the failure of the backup version of the effect, we get the needed dependence. The only thing one might call a failure in the backup process *preceding* (the backup version of) the effect is the nonoccurrence of N being present in a resting state at $t_2$. But hold fixed this nonoccurrence and it is far from clear you get the needed result: "If $c$ hadn't occurred and (still) N's resting presence at $t_2$ had not occurred, then $e$ wouldn't have occurred"—this seems false, because it seems that in the antecedent situation, N might be firing at $t_2$. ("Might"? Well, someone may suggest that the nonoccurrence of N's resting presence at $t_2$ might hold in virtue of N being absent *altogether* at $t_2$, in which case $e$ does not occur. I am willing to entertain that this *might*

be how that nonoccurrence is realized, but that this is what *would* happen seems very artificial.) Now, Hitchcock requires intermediate variables on a route to be "distinct" from the other variables on that route, by which he means, in the binary case, that the events corresponding to those variables are essentially distinct.[10] Hence the nonoccurrence of the late version of the effect could not be held fixed, as it is never intermediate between cause $C$ and effect $E$.[11] Perhaps we could alter the definition of "intermediate" to fix this.[12] But notice that now the intuitive picture is sullied. The intuitive idea was supposed to be that we gain dependence by holding fixed (non)occurrences of causal-route intermediates between $c$ and $e$; but a potential version of the effect is not naturally taken to be such an intermediate.

There may also be a need to hold fixed the nonoccurrence of a backup *cause* that is not an intermediate between cause and effect. There is a worry (e.g., Hall 1997) that sometimes were the cause not to occur, some of the events that might occur in its place are ones that would be followed by the effect. For example, suppose a lone dart thrower also had a knife and a crochet needle at hand. It seems reasonable to think, "The dart flight was a cause of the balloon pop, though if the dart flight hadn't occurred, a knife flight might have occurred (causing the pop anyway)." Clearly, if the dart flight hadn't occurred, and still the knife and needle flights hadn't, the pop wouldn't have—so the

---

[10] If instead he meant *actually* distinct, then he would have systemic troubles due to the fact that every nonoccurrent event is actually distinct from itself (since it does not actually have any parts).

[11] And clearly it is not something we would intuitively count as an intermediate on a causal route to $E$. Hitchcock has confirmed (in correspondence) that he does not intend to allow holding fixed the nonoccurrence of a version of the effect.

[12] We may also need to narrow the interpretation of one of the elements of appropriateness for a model, that equations represent counterfactual relationships between "distinct" events only. I presume that, like most writers, by "distinct" he means *essentially* distinct. But it would need to mean *actually*. What ripple effects this blanket liberalization might have on the analysis I am not sure.

situational-dependence analysis readily yields the correct verdict. On Hitchcock's account, the knife and needle flights cannot be held fixed, because they are not intermediate between cause and effect: *intuitively* they are not intermediate, but also formally (the requisite equations are not there, since the equations only encode "would"-dependence and we only have that the knife (needle) flight *might* have occurred without the dart flight). So it seems Hitchcock's analysis and the intuitive picture it serves may be too narrow to be adequate.

Here is another illustration of the wider applicability of situational dependence. As we have seen in earlier chapters, in indeterministic settings sometimes a backup process just fails on its own, not as a result of the cause (and so there is no preemption). For example, picture two slightly radioactive samples, A and B, placed before a Geiger counter; the counter clicks once within the next minute; the placement of A caused the click, since A emitted a particle; even though we cannot say that the click depended on A's placement, we *can* say that, given the lack of emissions from B, without A's placement the click wouldn't have occurred. Here there are not two routes from A's placement to the click, but rather the processes are independent. In these two illustrations of the larger usefulness of situational dependence, we see we can hold fixed the failures of other potential processes *whether or not* those failures are effects of the cause; this, I think, is what makes it more useful.

## 5.6 Stephen Yablo's Analysis

Stephen Yablo (2002) proposes to equate event causation with a relation he calls "de facto dependence." Let's begin with a few background definitions. For any fact $G$, $G$ is a *cause maker* iff, for some $c$ and $e$, had $c$ not occurred but $G$ still held, $e$ would not have

occurred. "One can see how it would create a presumption in favor of *c*'s causing *e* that *e* depends on it modulo a factual condition *G*" (p. 134).

From the time of some event *c*, the way things would have gone had *c* not occurred is the *fallback scenario*, and the way they actually went is the *actual scenario*.

*e*'s *actual needs* given *G* are the events x in the actual scenario such that had x not occurred but *G* still held, *e* would not have occurred. The intuitive idea here is that *G* is a circumstance that puts *e* in need of x.

*e*'s *fallback needs* with respect to *c* are the events in the fallback scenario that *e* would have (outright) depended on had *c* not occurred. Fallback needs are not relative to a cause maker.

From what we have seen in previous Sections, it is clear that threat/savior events and switches are the final effect's actual needs for some *G*. But Yablo does not want to count threat/savior events as causes; and there are some switches he does not want to count (he does want to count others, as we will see). These events, he points out, make *themselves* needed: the switch prevents one process, which is the only reason it is "needed" (to deliver the other process); the threat/savior is only needed because of the threat that it itself produced. The events are responsible for altering the circumstances such that they become essential, though they would not otherwise be. Thus the "need" for them is "artificial." This notion of artificiality gets defined as follows.

Suppose *c* is an actual need of *e* (given some *G*). Fact *H* makes the need for *c* *artificial* iff *e*'s fallback needs with respect to *c* coincide with a *c*-free subset of *e*'s actual needs given *H*. (*H* may be *G*, but it need not.) The idea is that what *e* would have needed without *c* are all still actually needed—and *c* itself does not actually meet any of those

needs (more on "meeting needs" later). In the king example, had the addition of poison (*c*) not occurred but still there was poison in the drink later (*G*), the king's breath (*e*) would not have occurred. So *c* is an actual need of *e* (given *G*). Is the need for *c* artificial? Had *c* not occurred, *e* would have depended on the king's earlier breaths, on his earlier meals, on his having defeated foes in battle last month,.... These are the fallback needs with respect to *c*. But *e* still depends on these very events in actuality, so where *H* is, say, the fact that 2 = 2, *e*'s actual needs given *H* are *e*'s fallback needs with respect to *c*. Thus *H* makes the need for *c* artificial (p. 136).

The fallback needs with respect to an event can never contain that event, because by definition that event does not occur in the fallback scenario; so if *e*'s fallback needs with respect to *c* coincide with a subset of the actual needs given *H*, don't they coincide with a "*c*-free" subset? Well, a fallback need might have actual-need *c* as a "counterpart." Yablo explains that subsets "coinciding" can be a matter of their members having "counterparts" in the other set, rather than identity. Counterparts are events that meet the same need. They are to be paired off "in ways that preserve salient features of the case: energy expended, distance traveled, time taken, place in the larger structure of needs" (p. 137). He gives the example of someone who needs to consume $H_2O$ and actually gets it via Italian ice but would (in the fallback scenario) have gotten it via a glass of water—these two ingestions may be different events, but they meet the same need (one coincides with the other). ("Need" is never explicitly defined.)

The strategy now is to count as causes events that meet needs not exposed as artificial by any *H* that satisfies a certain condition. Specifically, if the only *H*s that make a need artificial are facts less "natural" than all the *G*s given which *c* is a need, then *c* is a cause.

(The concept of naturalness is left intuitive for the paper.[13]) One event *de facto depends* on another iff it is put in need of the other by a *G* more natural than any *H*s exposing the need as artificial (p. 138). *Causation* is de facto dependence between distinct, occurrent events.

Where is the danger that an *H* might make the need for a clear cause come out artificial, but the cause is saved by the fact that the *H* is relatively unnatural? Yablo does not suggest an example, and I have not thought of one.

Why allow *any* unnatural-seeming facts? Yablo wants to be able to count non-preemptive overdeterminers as causes, to count some switchings as causes, to count trumping events as causes, and to count events with very extrinsic occurrence conditions as effects (not that he refers to them as having extrinsic occurrence conditions, but they do; his example (p. 142) is like "Paul's bullet" (Section 3.4.2.2.2).) For this, he needs flexibility in what kind of fact *G* or *H* can be—as we will see, he relies on facts involving negative extrinsic properties of events, which seem pretty unnatural. (This is *my* answer to the question. I cannot say for sure what Yablo's answer would be.)

Yablo applies the theory to preemption using the example of two rock-throwers. Suzy's rock breaks the window and Billy's then sails through the empty frame. If Suzy's throw (*c*) had not occurred and still Billy's rock did not touch the window (*G*), then the shattering (*e*) would not have occurred. *G* puts *e* in need of *c*. Is the need artificial? If *c* had not occurred, *e* would have depended on Billy's throw, his rock traveling toward the window, etc. "These [needs] would seem to recur in the actual situation as needs for Suzy's throw, her rock's traveling toward the window, and so on. But then Suzy's throw

[13] Yablo says he is following Lewis in taking naturalness seriously as a metaphysical tool. See Lewis 1983.

meets the same need as was met in the fallback scenario by Billy's throw. The need was thus preexisting; it would still have had to be met even if Suzy had elected not to throw" (p. 138). I take it that the point here is that the fallback needs do not coincide with a *c-free* subset of the actual needs given the above *G*, since Billy's throw coincides only with Suzy's throw. But this is only to ask whether *G* itself makes the need for *c* artificial; Yablo does not explore whether some other circumstance might do so. I do not see any; probably if one does, it will be trickier, and less natural, than *G*.

Yablo does not raise the question of whether Billy's preempted throw will come out a cause on the analysis. But it appears it will not. Billy's throw appears to be an actual need only given a rather unnatural-seeming *G*: if it (*c*) hadn't occurred, and still Suzy's throw did not occur unaccompanied (*G*), the shattering wouldn't have occurred. Is there a more or equally natural *H* that makes the need artificial? The shattering's fallback needs with respect to *c* are Suzy's throw, the journey of her rock, the earlier presence of glass, etc. But where *H* is that Billy's rock does not touch the window, we saw that the shattering depends on each of these events given *H*. So the fallback needs coincide with a *c*-free subset of the actual needs given *H*—*H* makes the need for *c* artificial. *H* looks more natural than *G* (since *G* is negative, I would say).

We are supposed to be able to count standard non-preemptive overdeterminers as causes in the following way. Let the example be one in which both rocks hit the window at the same time. The shattering depends on Suzy's throw given the fact $G_S$ that Billy's rock does not hit the window unaccompanied, and on Billy's throw given the fact $G_B$ that Suzy's rock does not hit unaccompanied (pp. 139-40). Is the need for, say, Billy's throw artificial? The fallback needs are Suzy's throw, the flight of her rock, the earlier presence

of glass, etc. We can use $G_S$ to make Billy's throw artificial: given that Billy's rock does not hit alone ($G_S$), the shattering is dependent on each of the fallback needs. $G_B$ makes Billy's throw a need, but $G_S$ makes it artificial. Clearly we have a tie for naturalness. But then according to the definition of de facto dependence, the shattering *fails* to de facto depend on Billy's throw. Yablo says that "more natural than" is open to interpretation: sometimes we may have in mind "*strictly* more natural," other times "at least as natural as" (p. 140). Thus we *can* count non-preemptive overdeterminers as causes, though we waver in accordance with our wavering between interpretations. (I will have a comment on this later.)

Again by way of negative extrinsic predicates, Yablo counts trumping events as causes. We have the example where the prince's fate corresponds to the content of the first spell (Schaffer 2000b). Merlin casts a frog spell, then Morgana casts a frog spell. Holding fixed that no one who is not Merlin casts a first spell, if Merlin's spell hadn't occurred, the prince would not have become a frog. No equally or more natural $H$ exposing the need as artificial presents itself, says Yablo (though he admits he cannot rule it out). But what if $H$ is that Merlin's spell does not occur unaccompanied? Given $H$, the effect depends on Morgana's spell and on the earlier presence of the prince; these are the fallback needs with respect to Merlin's spell; so $H$ makes the need artificial. I don't see that $H$ is less natural than $G$. If they are equally natural, then according to Yablo we should waver, as we do for clear cases of non-preemptive overdetermination. Also, it seems to me that Morgana's spell is no worse a candidate for cause on this theory. It is an actual need given $H$, and I see no more natural fact that makes this need artificial: the fallback needs are Merlin's spell and the earlier presence of the prince;

given that Morgana's spell does not occur alone, these are still needed, but this given does not appear to be relatively natural. Rather, it again looks like a tie. (Yablo does not consider whether his theory will count Morgana's spell a cause.)

The analysis is supposed to count switches that make a "big" difference as causes, but not other switches. The idea is that the train journey along the branch not taken is a fallback need that is still needed in the actual scenario, because the journey on the actual branch is a counterpart of it; though if there were a big difference between the two branches, this counterpart relation would not hold, and the fallback needs would not all be subsumed by actual ones. Yablo points to Hall's example of "Billy's Kiss" (Hall 2000, p. 209): Suzy kisses Billy, so Billy takes the scenic route home instead of doing his errands; he whistles a happy tune; the whistling would have occurred anyway without the kiss, because Billy would have heard that tune in the store and it would have stuck in his head. We want to count the kiss a cause of the whistling. A fallback need is hearing the tune in the store; this does not occur as an actual need because it does not actually occur and no counterpart of it occurs either.

This attempt to count only some switchings seems strained, to me. Suppose that without the kiss, Billy would have thought about his great new job, walked a different route and then whistled as a result of these happy job-thoughts. It still seems to me that, since he actually whistled as a result of happy girl-thoughts, the kiss was a cause of his whistling; but it also seems that the two happy-thought strolls are counterparts, that they meet the same need. So it seems Yablo has not found the correct dividing line.

As for its causal verdicts, I conclude that the theory is no improvement over others with respect to trumping and that it does not appear to achieve what it sets out to achieve

with respect to switching. With respect to threat/savior cases, some will like Yablo's results; but I, as I said, prefer to count those cases as causation. The standard non-preemptive overdetermination result is good.

But I must say I find the motivation for this last result intuitively alien. It seems to me that the fact that Suzy's throw does not occur alone is no part of why I think Billy's throw is a cause (nor even of why I might be thinking of Billy's throw as in some sense "needed").

More generally, it does not seem that we make comparisons involving negative extrinsic predicates in judging whether some event was a cause, and for this reason the theory lacks intuitive support. In judging that Lucy's (preempted) decision to throw was not a cause, did you entertain the fact that the only way it is needed is via a rather unnatural-seeming $G$ (e.g., that my throw did not occur unaccompanied)?

The theory gets results in threat/savior cases that some may find attractive. But to me the theory seems more complex, more subtle (due to "natural") and less intuitive than the theory I have proposed.

# CHAPTER 6

# TOWARD AN ANALYSIS FOR PURELY INDETERMINISTIC WORLDS

## 6.1 Introduction

Just as an analysis for deterministic worlds may be helpful in finding an analysis suitable for all worlds, so, too, may an analysis for purely *in*deterministic worlds—worlds where at any given time, any possible future pattern of events has some non-zero chance. In addition to being stepping stones, analyses for these two extremes are of some interest in themselves: they may analyze species of our causal concept that correspond to ways we habitually think about the world—sometimes as deterministic, sometimes as purely indeterministic. Despite the great attention that deterministic analyses have received, an analysis for purely indeterministic worlds is, I think, in line with the more common way people naturally think about the world: nothing really has *no* probability; the freaky and miraculous are just *very* improbable. In this chapter, I offer an analysis of causation for purely indeterministic worlds. In contrast with the deterministic analysis of Chapter 5, my proposal here is rather inelegant and, as we will see, at points not fully motivated. Hence the title of this chapter.

## 6.2 Obstacles to Be Overcome

Very roughly, the idea behind the analysis is that causes are connected to effects by a chain of (a kind of) direct causation, where "nonoccurrences" may be intermediate links; this direct causation is primarily a matter of making the direct effect more probable. Let us first consider the difficulties of this general approach, and get a hint of how they can be solved. There are six to consider, the first two of which we have already encountered.

#1) We saw that a preempting cause may make the effect *less* probable by threatening to prevent a much more reliable alternate process to the effect. This can happen even in cases of direct causation: the whole unreliable-but-successful process up to the effect is a direct cause without which the effect might have been equally or more probable. For example, in the dart-throw preemption case from previous chapters, if Lucy's dart is sharp and mine is dull, the chance of the balloon-pop might have been greater without my dart flight (at the time at the end of the flight), because Lucy's sharper dart might have been in play instead of mine.

I have pressed, however, that preemption always involves some failure in the backup process (here, a missing dart flight). We will see that my analysis will take advantage of these failures when assessing probabilities. The basic idea will be, *given* the incompleteness in the backup process (Lucy's dart flight), the successful process *does* make the effect more probable.

#2) We also saw the problem of failed potential causes. Sometimes a potential causal process leading to the effect fails to complete, so that it, and its parts, make the effect more probable but do not cause it. Such a would-be cause can occur simultaneously with a direct cause, so that there is no hope of locating genuine direct causes just by the times of their chance-contributions. We saw this in Figure 4.1: the non-cause $c$ makes $e$ more probable, and $c$ occurs simultaneously with $b$. We saw it also in overlapping (Section 3.4.1).

But failed potential causes of an event $e$ are failed in virtue of failing to cause other things, be it potential causal intermediates—$f$, in Figure 4.1—or, in the case of overlapping, other effects besides $e$—the queen turning into a frog, in the overlapping

example. Again, we can hold fixed these failures, and the troublesome chance-contributions go away: *given* the nonoccurrence of the intermediate (*f*), *c* does *not* make *e* more probable; given the nonoccurrence of the queen's transformation, spell PQ does not make the prince's transformation more probable. (We will look at this more closely later.)

#3) Here is an example we could count as a special kind of overlapping. Suppose that without spell G or spell H, that the sad prince will become happy, turn green or become a frog each has an independent chance of .01. If spell G is cast alone, the chance of the prince becoming a green frog is .2, while the chance of him becoming happy is unchanged; if spell H is cast alone, the chance of the prince becoming a happy frog is .2, while the chance of him becoming green is unchanged. Neither spell improves the chance that only one element of its corresponding pair (green/frog, happy/frog) will be realized—it contributes only to the chance of the pair. The occurrence of each pair is, in any situation, probabilistically independent of the occurrence of the other pair's elements and of the occurrence of the other kind of spell—modulo the independent background chances (for example, with no spells, the chance of a happy frog is .0001; add G, and it must go up, to (.2 x .01) + (.8 x .0001), or .00208). Hence if both spells are cast, the chance of the prince becoming a frog (be it green, happy, or both) is .36. Today, both spells are cast, and the prince becomes a green, but still sad, frog. It seems that, since H contributes only to the chance of the frog/happy *pair* (and what it contributes to the other pair is only a byproduct of contributing to the frog/happy pair), it did not help, or was not a cause of, the prince becoming a frog, though it made it more probable. (We can view

this as a kind of condensed overlapping, where the pair of effects associated with each spell is a pair that essentially occurs in one region.)

Again, this challenge will be met by holding fixed a certain failure: *given* that happiness will not occur, H does *not* make the chance of the prince becoming a frog more probable.

For the difficulties that follow, picture a simple causal setting, where there are not competing processes leading to the effect.

#4) Suppose *b*+*c* is an event that occurs as long as either *b* or *c* occurs, though in fact both *b* and *c* occur. Suppose *a* is a direct cause of *b* and makes *b* more probable, and *c* is a direct cause of *d* and makes *d* more probable (also, let *a* be probabilistically irrelevant to *c*, and let *b* be probabilistically irrelevant to *d*). Then *a* will make *b*+*c* more probable, and *b*+*c* will make *d* more probable (because in supposing away *b*+*c*, you have to suppose away *c*); so *a* will come out a cause of *d*, even if it is not.

This problem is solved by requiring that an event contribute to the chance of every sufficient part of *e* in order to be a direct cause of *e*. *a* does not contribute to the chance of *c*, which is a sufficient part of *b*+*c*. The analysis I will propose is built upon this requirement of chance-contribution to every sufficient part.

#5) In a purely chancy world, it seems that, for any event, had it not occurred, a great variety of other events *might* have occurred in its place, since they all would have had some non-zero chance. But then it may be that, without the direct cause *d*, the direct effect might have been no less probable, thanks to one of these replacements.

Notice that the event consisting in the union of *d* and all such nuisance replacements is an event such that, without it, the direct effect would have been less probable. So there is

no replacements problem for this weak direct cause. Since *d* is a non-proper part (Section 2.4) of this weak cause, we can count *d* by allowing that every non-proper part of a cause is also a cause. This is how the analysis will deal with this problem.

#6) Carl catches the ball that was headed for the window, so the catch is a cause of the later presence of intact window. I was spraying red paint in the space between the catch and the window. My spraying made the nonoccurrence of a blue rock-flight there more probable, and it seems that the nonoccurrence of a blue rock-flight there made the intact window presence more probable (because had a blue rock-flight occurred there, the presence would have been very improbable). Yet my spraying was not a cause of the window presence.

Notice that for the region between the catch and the window-presence there is a strongest nonoccurrence that makes the window presence more probable, namely the nonoccurrence of the event consisting in the union of events that would have made it less probable: if this union had occurred, the chance of the window presence would have been lower. This union includes projectile flights of various colors. While the spraying increases the chance of a red flight, it correspondingly decreases the chance of a non-red flight, and is therefore irrelevant to the chance of the strongest nonoccurrence. Carl's catch, though, makes this nonoccurrence more probable. Hence we can exclude the spraying by requiring that causes be connected to effects by chains of probability-raising that involve *strongest* nonoccurrences only.

My analysis is based on the intuition that direct causes are "directly" positively probabilistically relevant under the circumstances. In the next section, I will motivate and give a definition of this sort of relevance. The first three problems above are solved

via this special type of relevance. The analysis of causation will be built upon chains of direct causation in which direct causes are relevant in this way to the chance of every sufficient part of their direct effects (to handle problem #4) and the nonoccurrence links are strongest ones only (#6)—unfortunately I have no clear motivation for these two maneuvers, beyond their effectiveness. They do, however, seem to be ingredients in an intuitive notion of "primary" responsibility: it is not the absence of a blue rock flight that is primarily responsible for the window being intact, but rather the absence of anything that could have broken it; it is not the whole event *e* that *d* is primarily responsible for if there are sufficient parts of *e* (kind of like little *e*'s) that escape *d*'s help. Those chains of direct causation will be called chains of primary responsibility. Problem #5 is solved in a last move—I just add that non-proper parts of causes are causes, as I trust is intuitive enough.

6.3 Positive Relevance Under the Circumstances

Let *A* be a simple event proposition, either O(*a*) or ~O(*a*). Likewise for *B*, *C*, *D* and *E*. Where *A* is true, if *A* is O(*a*), then *A's region* is the region in which *a* occurs, and if *A* is ~O(*a*), then *A*'s region is the union of regions in which *a* could possibly occur.

I want to begin with some intuitive remarks about probabilistic relevance, setting aside any precise interpretations of the probabilistic expressions I use.

As I said, my analysis is based on the intuition that direct causes are "directly" positively probabilistically relevant under the circumstances. Sometimes an event is probabilistically relevant to another in virtue of its relevance to yet other events. For example, the gunshot makes the wound more probable in virtue of the fact that it makes the flight of a bullet more probable. This is reflected in the fact that the flight makes the

shot probabilistically irrelevant to the wound, or "screens off" (Reichenbach) the shot from the wound: the wound's probability given the flight and the shot is equal to its probability given the flight and no shot. We see this not only in indirect causes, but in indirect effects: the wound makes the shot more probable, though only because it makes a bullet flight more probable. And in forking effects: the bullet flight makes the sound of a gunshot more probable, but only because it makes a gunshot more probable. In every case, the "parasitic" nature of the probabilistic relevance is revealed by the fact that the relevant event can be screened by some actual (pattern of) events (that excludes the two events at issue). Perhaps we can define a kind of *non*-parasitic relevance, what we might call "direct relevance," in terms of an inability to be screened. Roughly, the idea would be to say $A$ is *directly relevant* to $B$ *given* $G$ (some true background conditions entailing nothing about the $A$ and $B$ regions) iff (i) $\Pr(B|G\&A) \neq \Pr(B|G\&{\sim}A)$, and (ii) for any conjunction $F$ of true event propositions (entailing nothing about the $A$ and $B$ regions), $\Pr(B|G\&F\&A) \neq \Pr(B|G\&F\&{\sim}A)$.

With respect to the intuition of probabilistic relevance under the circumstances, I shall interpret "the circumstances" very liberally. It will include the entire pattern of events (past, present and future) outside the (candidate) direct cause and effect. Call this pattern "$O$." Notice that if $A$ is relevant to $B$ given $O$, it is directly so: $A$ is not screenable from $B$ given $O$, since $O$ already entails every candidate conjunction $F$ of true event propositions. This relevance all-else-being-equal is a direct relevance. Therefore, once I define positive relevance under the circumstances, I will have cashed out the intuition I started with, that direct causes are "directly" positively relevant under the circumstances.

Chances hold "at a time" in virtue of the (actual) spatiotemporal region that comprises all of the past up through that time (Section 2.3). Ultimately it is the complete pattern of events in this region, and the laws, that combine to entail some chance. Presumably other patterns besides past ones up to a time can, with the laws, entail chances. Say that a precise *pattern* of events is a set of true simple event propositions, where each involved event is spatiotemporally inflexible. In addition to our chance function (Section 2.3) that takes a world, a time and an event proposition, I presume there is a chance function that takes a world, a pattern and an event proposition and delivers the correct chance of that proposition.

I presume that the entire pattern of events outside $B$'s region (with the laws) entails a chance of $B$; call this $B$'s *world-chance* (and "$b$'s world-chance" if $B$ is O($b$)). Say $A$ *contributes to* $B$'s *world-chance* iff, were $A$ not the case but everything outside $A$'s region were exactly the same, $B$'s world-chance would be lower than it actually is.

In addition to including $O$, the circumstances of $B$ will also include certain simple event propositions whose regions are within $B$'s. Recall from the previous section that a goal in holding things fixed is to assess probabilities in light of failed potential effects. As far as I can see, every contributor to an event's world-chance is a potential cause of it. In order to get the right results, we will need to hold fixed the failures of potential effects of a candidate direct cause, $A$; in holding fixed $O$, we get most of them, but we may also need to hold fixed some within $B$'s region. We can locate these potential effects by whether $A$ contributes to their world-chances. Suppose $A$ and $B$ are true and distinct. Let $S$ be the set of all true simple spatiotemporally-inflexible-event propositions that do not entail $B$. Construct pattern $P$ as follows. First, put in every member of $S$ whose region is

outside *B*'s actual region. Then, add any remaining failures of potential effects of *A*: add every remaining member such that *A* contributes to the world-chance of its negation. I presume that *P* (with the laws) will entail a chance of *B*—call this *B*'s *world-chance modulo A's failures*. If there are no failures of potential effects of *A* in *B*'s region, then the pattern that determines *B*'s world-chance modulo *A*'s failures is the same pattern that determines *B*'s world-chance.[1]

Where *A* and *B* are true and their regions distinct, *A is positively relevant to B under the circumstances* iff, were *A* not the case but everything outside *A*'s region were exactly the same, *B*'s world-chance modulo *A*'s failures would be lower than it actually is. For short: *A* makes a *circumstantial contribution* to *B*. Where *A* is O(*a*), or *B* is O(*b*), we can also substitute the event for the proposition in the appropriate place: *a* makes a circumstantial contribution to *B*, or to *b*, etc.

## 6.4 Primary Responsibility and Causation

The backbone of the analysis is a special chain of circumstantial contribution. I will refer to it as a chain of primary responsibility. By definition, "*A* is primarily responsible for *B*" means that (i) *A* and *B* are true, (ii) all of *A*'s region is earlier than *B*'s,[2] and (iii) the relevant following condition holds:

case 1: *A* is O(*a*), *B* is O(*b*): *A* makes a circumstantial contribution to every sufficient part of *b*.

---

[1] Thanks are due to Jonathan Schaffer for stimulating some of the refinements here.

[2] This temporal move is, I confess, pretty ad hoc. Assuming there is no such thing as causation at a distance, it allows the analysis to differentiate direct causes from direct effects and from direct effects of a common cause; but I am not sure what other motivation it has than that. (I shall assume there is no such thing as direct causation at a distance; as I said in Section 4.3, I am not confident that such causation is metaphysically possible.)

case 2: $A$ is ~O($a$), $B$ is O($b$): $A$ makes a circumstantial contribution to every sufficient part of $b$; this is not true for any strengthening of $A$.
case 3: $A$ is O($a$), $B$ is ~O($b$): $A$ makes a circumstantial contribution to $B$.
case 4: $A$ is ~O($a$), $B$ is ~O($b$): $A$ makes a circumstantial contribution to $B$; this is not true for any strengthening of $A$.

A chain of primary responsibility is a sequence of simple event propositions, each of which (but the last) is primarily responsible for the next. $c$ is *cause* of $e$ iff $c$ is a non-proper part of some event $c_1$ such that there is a chain of primary responsibility from $c_1$ to $e$.

A rock shatters a window. Let $c$ be the release of the rock, let $d_1$ and $d_2$ be the first and second halves of the rock's journey, and let $e$ be the shattering. $d_1$'s world-chance is high. Let us assume there are no significant failed potential effects of $A$ in $B$'s region. Had $c$ not occurred, but everything outside $c$'s region had remained unchanged, $d_1$'s world-chance might, let us say, have been even higher, because a more reliable release might have occurred in $c$'s place. Consider the event $f$ which is the union of $c$ and all such more reliable releases; had $f$ not occurred, but everything else outside the release's actual region had remained the same, *then* $d_1$'s world-chance would have been lower than it actually was. And it seems safe to say this about all of $d_1$'s sufficient parts. So $f$ makes a circumstantial contribution to every sufficient part of $d_1$. Since $c$ is a non-proper part of $f$, it comes out a cause of $d_1$. $d_1$ similarly makes a circumstantial contribution to every sufficient part of $d_2$, and $d_2$ to the shattering. Since $c$ is a non-proper part of $f$, it is a non-proper part of an event from which there is a chain of primary responsibility to $e$; hence it comes out a cause of $e$.

Notice that $c$ does not make a circumstantial contribution to (every sufficient part of) $d_2$. In supposing away $c$ but leaving $d_1$, $d_2$'s world-chance remains the same—because $d_1$ screens $c$ from $d_2$. Only *direct* causes make circumstantial contributions.

In cases of preemption, we have seen that the failed backup process is failed in virtue of some occurrence or nonoccurrence outside the effect's region. In assessing world-chances, this failure will always be held fixed. Given this failure, the direct cause in the successful process will make the effect more probable. For example, consider the two-rock case. The circumstances of the shattering include the lack of a late shattering. The shattering would be less likely in these circumstances if they did not include a last stage of the first rock throw (or any window-threatening event of which that stage is a non-proper part). Or consider the dart-throw case. The circumstances of the balloon-pop include the lack of a second dart. The pop would be less likely in these circumstances if they lacked the last stage of the first dart throw. The various diagrams of preemption in previous chapters represent some neurons that fail to fire; it is easy to see that, given these failures, the direct causes make the effects more probable in the relevant way. Indirect causes then get counted by way of intermediate ones. (Nonoccurrence links will fulfill the same function. We will look closely at some chains of that sort below.)

Occurrent preempted potential causes—or backups—will not be counted as causes. In order for the effect to occur by way of the unsuccessful process, the aforementioned failures need to occur—and they are either causally intermediate between the backups and the effect or they are the version of the effect that marks the end of that process (e.g., the late shattering). In its circumstances, which include the failure(s), the effect is no more probable with the backup than without. For example, taking into account the lack

of the late shattering, a late stage of the second rock throw does not make the (flexible) shattering more probable in its circumstances.

Now consider the problem of failed potential causes. We have the type of case where a potential indirect cause *a* makes the effect more probable at *a*'s time, but *a* is still not a cause: the presence of a radioactive sample makes the Geiger counter's "click" more probable, but the click is caused by a different sample (Section 3.4.1). This problem evaporates, because the inactive sample is only a potential *indirect* cause—the click is not more probable in its circumstances thanks to the samples, because the flight of the particle from one sample and the lack of such a flight from the other screen the samples from the click.

Then there is the overlapping type of case. Here is Jonathan Schaffer's example again (as adapted for pure indeterminism). One wizard casts spell PK, which is of a type that generates a .5 chance of the prince and the king becoming frogs, while another casts PQ, which is of a type that generates a .5 chance of the prince and the queen becoming frogs (these chances are the chances at the time of spells); the occurrence of each transformation *pair* is, in any situation, probabilistically independent of the occurrence of the other and of the occurrence of the other type of spell, modulo the independent background chances; and neither spell contributes to the chance that *only one* of its potential victims becomes a frog. So the prince's chance of becoming a frog is .75. As it turns out, the king and prince, but not the queen, turn into frogs. So it seems PK was a cause of the prince becoming a frog, even though both spells made it more probable at the time they were cast.

But PQ does not make a circumstantial contribution to the prince's transformation. The world-chance of that transformation is the same with or without PQ, because it holds fixed that the queen does not turn; given that she does not turn, PQ does not make the prince's transformation more probable. (Think of it in terms of evidence. You don't know whether the prince turned, and you're wondering the probability of it. At first you know only that PK occurred and that the king turned. Then you learn that the queen did not turn. You now know that even if PQ occurred, it did not "help" anything; if the prince turned, it was either independently or thanks to PK; to learn now that PQ occurred does not increase your confidence in the proposition that the prince turned.)[3]

Consider now the condensed overlapping case (problem #3), the story of the sad green frog. Spell H contributes to the world-chance of a happy frog; but the frog was not happy. Here the pattern that determines the world-chance of the effect modulo H's failures includes the nonoccurrence of a happy frog. Given the effect's surroundings, and also the lack of a happy frog, the chance of there being a frog is the same with or without H. The case is analogous to that of the previous paragraph.

The analysis implies that a direct cause makes a circumstantial contribution to every sufficient part of the effect, that a non-proper part of a cause is also a cause, and that nonoccurrences involved in chains of primary responsibility are only the strongest version that will serve. These were what were needed to overcome problems #4-6. Let's look closely at #6.

---

[3] Does PK count as a cause? One might think that it does not because given the king's transformation, the world-chance of the prince's is 1 with or without A. But we are restricting ourselves to purely indeterministic worlds, wherein every pattern of events is legal. Thus no world-chance is 1. There is some chance of the king turning without the prince doing so, whether PK occurs or not, though if PK occurs this is less likely. PK contributes to the effect's world-chance: the effect's circumstances would, if they lacked PK, make the effect less likely.

The spraying of red paint apparently raised the chance (at its time) of there being no blue rock flight up to the window, which nonoccurrence in turn raised the chance of the (intact) window-presence, though truly the spraying is not a cause of the window-presence. We want the catch that saved the window to come out a cause, but not the spraying. Consider the surrounding circumstances of the window-presence. If they included a blue rock flight up to the window, they would make (every sufficient part of) the window-presence less probable, so the nonoccurrence of such a flight does makes a circumstantial contribution (the nonoccurrence of a blue rock-flight does not contribute to the world-chances of any window-region goings on that failed). However, in order for this nonoccurrence to be primarily responsible for the window-presence, there needs to be no strengthening of it that makes such a contribution—yet one does. For there is a strongest nonoccurrence—the nonoccurrence of all the events conducive of window-breaking—such that the window-presence's chance would be lower without *this* nonoccurrence (that is to say, *with* one of those events). If the spraying is to come out a cause, it will have to be primarily responsible for this strong nonoccurrence—it will have to make a circumstantial contribution to it (case 3). It appears not to, however. Here we have to hold fixed the lack of a red flight, since the spraying contributes to the world-chance of such a flight. Given the surroundings and the lack of a red flight, the chance of the strong window-threatening nonoccurrence is no higher with the spraying than without. The catch, on the other hand, *does* contribute to the chance of it (here there is no need to hold fixed things in *B*'s region).

We saw the following challenge to transitivity earlier (Section 4.9): adding green food-coloring to the poisoned drink caused the drinking of a green, poisoned drink, and

this drinking in turn caused a death, but the addition of food-coloring was not a cause of that death. The challenge carries over to the probabilistic approach: on the face of it, the addition of food-coloring makes that drinking of a green, poisoned drink more probable (at the time of the addition), and that drinking makes the death more probable. Let's make sure my proposal is not vulnerable to this attack.

If the (essentially) green drinking had not occurred, other deadly events might have occurred in its place, such as yellow drinkings, scalding-hot drinkings, .... But there is a weakening of the green drinking that includes such alternates and without which the world-chance of the death would have been lower (throughout this example, we do not need to hold fixed anything in *B*'s region). Call this event *d*. It is *d* that makes a circumstantial contribution, not the green drinking. For the addition of food-coloring to be a cause, it has to be (a non-proper part of) an event that makes a circumstantial contribution to every sufficient part of *d*. But it is not: it does not make one to the drinking that is only accidentally green.

## 6.5 Regrets

As I said, in my opinion the analysis is less elegant than we should like, and certain maneuvers are not adequately motivated. There may be a further shortcoming. In Section 5.4.1 we saw this example: A safe is falling towards a doughnut on the sidewalk; I bump the doughnut left a foot, so Squash occurs in region S instead of R. Let R-squash and S-squash be squashes that essentially occur in R and S, respectively. Squash, the effect at issue, is more flexible and could occur in either region. If the world is deterministic, then we might say this: (Q) "If the bump hadn't occurred but still R-squash hadn't occurred, Squash wouldn't have occurred." If Q is true, then the bump is a

cause of Squash, since it is situationally dependent on it. I pointed out, however, that there is flexibility in the situational dependence analysis, since we might not want to accept Q: we might take it that, in the antecedent situation, Squash occurring somewhere besides R (say, in S) is no more far-fetched than its failing to occur at all. This sort of flexibility is lacking in the purely indeterministic analysis, because the bump contributes to the world-chance of Squash (given the lack of R-squash in the circumstances, Squash is more probable with the bump than without). This lack of flexibility is a result of how precisely the analysis dictates the nature of its counterfactual comparisons. Perhaps the analysis is not as flexible as it ought to be.

# BIBLIOGRAPHY


Anscombe, G. E. M. (1971): *Causality and Determination: An Inaugural Lecture*. Cambridge: Cambridge University Press.

Armstrong, D. M. (1997): *A World of States of Affairs*. Cambridge: Cambridge University Press.

——— (1999): "The Open Door: Counterfactual vs. Singularist Theories of Causation," in H. Sankey (ed.), *Causation and Laws of Nature* (Boston: Kluwer Academic).

Beebee, Helen (1998): "Do Causes Raise the Chances of Effects," *Analysis*, 58, pp. 182–90.

Bennett, Jonathan (1984): "Counterfactuals and Temporal Direction," *The Philosophical Review*, 93, pp. 57–91.

——— (1987): "Event Causation: The Counterfactual Analysis," *Philosophical Perspectives*, 1, pp. 367–86.

——— (1988): *Events and Their Names*. Indianapolis: Hackett Publishing.

Bunzl, Martin (1979): "Causal Overdetermination," *Journal of Philosophy*, 76, pp. 134–50.

Byrne, Alex and Hall, Ned (1998): "Against the PCA-Analysis," *Analysis*, 58 pp. 38–44.

Cartwright, Nancy (1979): "Causal Laws and Effective Strategies," *Noûs*, 13, pp. 419–37.

Collins, John (2000): "Preemptive Prevention," *Journal of Philosophy*, 97, pp. 223–34.

Davidson, Donald (1967): "Causal Relations," in Davidson 1980, pp. 149–62. Originally in *Journal of Philosophy*, 64.

——— (1969): "The Individuation of Events", in Davidson 1980, pp. 163–80. Originally in Nicholas Rescher (ed.), *Essays in Honor of Carl G. Hempel* (Dordrecht: D. Reidel Publishing).

——— (1980): *Essays on Actions and Events*. Oxford: Clarendon Press.

Dowe, Phil (1992): "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory," *Philosophy of Science*, 59, pp. 195–216.

——— (1995): "Causality and Conserved Quantities: A Reply to Salmon," *Philosophy of Science*, 62, pp. 321–33.

——— (2001): "A Counterfactual Theory of Prevention and 'Causation' by Omission," *Australasian Journal of Philosophy*, 79, pp. 216–26.

Ducasse, C. J. (1926): "On the Nature and the Observability of the Causal Relation," *Journal of Philosophy*, 23, pp. 57–68.

Ehring, Douglas (1997): *Causation and Persistence: A Theory of Causation*. Oxford: Oxford University Press.

Fair, David (1979): "Causation and the Flow of Energy," *Erkenntnis*, 14, pp. 219–50.

Frankfurt, Harry (1969): "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy*, 66, pp. 829–39.

Ganeri, Jonardon, Noordhof, Paul and Ramachandran, Murali (1996): "Counterfactuals and Preemptive Causation," *Analysis*, 56, pp. 216–25.

——— (1998): "For a (revised) PCA-analysis," *Analysis*, 58, pp. 45–7.

Goldman, Alvin I. (1970): *A Theory of Human Action*. Englewood Cliffs, NJ: Prentice-Hall.

——— (1977): "Perceptual Objects," *Synthese*, 35, pp. 257–84.

Goodman, Nelson (1955): *Fact, Fiction, and Forecast*. Cambridge, Mass.: Harvard University Press.

Goosens, William K. (1979): "Causal Chains and Counterfactuals," *Journal of Philosophy*, 76, pp. 489–95.

Hall, Ned (1997): "Two Concepts of Causation" (unpublished manuscript).

——— (1994): "Correcting the Guide to Objective Chance," *Mind*, 103, pp. 505–17.

——— (2000): "Causation and the Price of Transitivity," *Journal of Philosophy*, 97, pp. 198–222.

Hausman, Daniel M. (1998): *Causal Asymmetries*. Cambridge: Cambridge University Press.

Hempel, Carl G. (1965): *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.

Hitchcock, Christopher (2001): "The Intransitivity of Causation Revealed in Equations and Graphs," *Journal of Philosophy*, 98, pp. 273–99.

Hume, David (1739): *A Treatise of Human Nature*, 2d ed., edited by L. A. Selby-Bigge and P. H. Nidditch (1978). Oxford: Oxford University Press.

——— (1748): *Enquiry Concerning Human Understanding*, edited by C. W. Hendel (1955). Indianapolis: Bobbs-Merrill Company.

Kim, Jaegwon (1973a): "Causes and Counterfactuals," in Sosa and Tooley 1993, pp. 205–7. Originally in *Journal of Philosophy*, 70, 1973.

——— (1973b): "Causation, Nomic Subsumption and the Concept of Event," in Kim 1993, pp. 3–21. Originally in *Journal of Philosophy*, 70, 1973.

——— (1976): "Events as Property Exemplifications," in Kim 1993, pp. 33–52.

——— (1993): *Supervenience and Mind*. Cambridge: Cambridge University Press.

Krasner, Daniel and Heller, Mark (1994): "The Miracle of Counterfactuals: Counterexamples to Lewis's World Ordering," *Philosophical Studies*, 76, pp. 27–43.

Kvart, Igal (1986): *A Theory of Counterfactuals*. Indianapolis: Hackett Publishing Co.

——— (1997): "Cause and Some Positive Causal Impact," in James Tomberlin (ed.), *Philosophical Perspectives 11: Mind, Causation, and World, 1997* (Malden, Mass.: Blackwell Publishers).

——— (2001): "Lewis's 'Causation as Influence'," *Australasian Journal of Philosophy*, 79, pp. 409–21.

Lee, Jig-chuen (1988): "The Nontransitivity of Causation," *American Philosophical Quarterly*, 25, pp. 87–94.

Lewis, David (1970): "How to Define Theoretical Terms," *Journal of Philosophy*, 67, pp. 427–46.

——— (1973a): *Counterfactuals*. Cambridge, Mass.: Harvard University Press.

——— (1973b): "Causation," in Lewis 1986a, pp. 159–72. Originally in *Journal of Philosophy*, 70, 1973.

——— (1979): "Counterfactual Dependence and Time's Arrow," in Lewis 1986a, pp. 32–66. Originally in *Noûs*, 13, 1979.

——— (1980): "A Subjectivist's Guide to Objective Chance," in Lewis 1986a, pp. 83–113. Originally in Richard C. Jeffrey (ed.), *Studies in Inductive Logic and Probability, Vol. II* (Berkeley: University of California).

——— (1983): "New Work for a Theory of Universals," *Australasian Journal of Philosophy*, 61, pp. 343–77.

——— (1986a): *Philosophical Papers, Vol. II*. Oxford: Oxford University Press.

——— (1986b): "Events," in Lewis 1986a, pp. 241–69.

——— (1986c): "Causal Explanation," in Lewis 1986a, pp. 214–40.

——— (1986d): Postscripts to "Causation," in Lewis 1986a, pp. 172–213.

——— (1991): *Parts of Classes*. Oxford: Basil Blackwell.

——— (1994): "Humean Supervenience Debugged," *Mind*, 103, pp. 473–90.

——— (2000): "Causation as Influence," *Journal of Philosophy*, 97, pp. 182–97.

Lyon, Ardon (1967): "Causality," *British Journal for the Philosophy of Science*, 18, pp. 1–20.

Mackie, John L. (1965): "Causes and Conditions," *American Philosophical Quarterly*, 2, pp. 245–64.

——— (1974): *The Cement of the Universe*. Oxford: Oxford University Press.

Mackie, Penelope (1992): "Causing, Delaying, and Hastening: Do Rains Cause Fires?" *Mind*, 101, pp. 323–44.

McDermott, Michael (1995a): "Lewis on Causal Dependence," *Australasian Journal of Philosophy*, 73, pp. 129–39.

——— (1995b): "Redundant Causation," *British Journal for the Philosophy of Science*, 46, pp. 523–44.

——— (1997): "Metaphysics and Conceptual Analysis: Lewis on Indeterministic Causation," *Australasian Journal of Philosophy*, 75, pp. 396–403.

——— (2002): "Causation: Influence versus Sufficiency," *Journal of Philosophy*, 99, pp. 84–101.

Mellor, D. H. (1995): *The Facts of Causation*. London: Routledge.

Menzies, Peter (1989a): "Probabilistic Causation and Causal Processes: a Critique of Lewis," *Philosophy of Science*, 56, pp. 642–63.

——— (1989b): "A Unified Account of Causal Relata," *Australasian Journal of Philosophy*, 67, pp. 59–83.

——— (1996): "Probabilistic Causation and the Pre-emption Problem," *Mind*, 105, pp. 85–116.

Menzies, Peter and Price, Huw (1993): "Causation as a Secondary Quality," *British Journal for the Philosophy of Science*, 44, pp. 187–203.

Noordhof, Paul (1998): "Problems for the M-Set Analysis of Causation," *Mind*, 107, pp. 457–63.

——— (1999): "Probabilistic Causation, Preemption and Counterfactuals," *Mind*, 108, pp. 95–125.

——— (2000): "Ramachandran's Four Counterexamples," *Mind*, 109, pp. 315–24.

Paul, L. A. (1998a): "Problems with Late Preemption," *Analysis*, 58, pp. 48–53.

——— (1998b): "Keeping Track of the Time," *Analysis*, 58, pp. 191–8.

——— (2000): "Aspect Causation," *Journal of Philosophy*, 97, pp. 223–34.

Pearl, Judea (2000): *Causality*. Cambridge: Cambridge University Press.

Quine, Willard Van Orman (1960): *Word and Object*. Cambridge, Mass.: MIT Press.

Ramachandran, Murali (1997): "A Counterfactual Analysis of Causation," *Mind*, 106, pp. 264–77.

——— (1998): "The M-Set Analysis of Causation: Objections and Responses," *Mind*, 107, pp. 465–71.

——— (2000): "Noordhof on Probabilistic Causation," *Mind*, 109, pp. 309–13.

Reichenbach, Hans (1956): *The Direction of Time*. Berkeley and Los Angeles: University of California Press.

Russell, Bertrand (1912–13): "On the Notion of Cause," *Proceedings of the Aristotelian Society*, 13, pp. 1–26.

Salmon, Wesley C. (1980a): "Probabilistic Causality," *Pacific Philosophical Quarterly*, 61, pp. 50–74.

——— (1980b): "Causality: Production and Propagation," in P. D. Asquith and R. N. Giere (eds), *Philosophy of Science Association 1980, Vol. 2*, pp. 49–69.

——— (1994): "Causality without Counterfactuals," *Philosophy of Science*, 61, pp. 297–312.

Schaffer, Jonathan (2000a): "Overlappings: Probability-Raising without Causation," *Australasian Journal of Philosophy*, 78, pp. 40–6.

——— (2000b): "Trumping Preemption," *Journal of Philosophy*, 97, pp. 165–81.

——— (2000c): "Causation by Disconnection," *Philosophy of Science*, 67, pp. 285–300.

——— (2001a): "Causation, Influence and Effluence," *Analysis*, 61, pp. 11–18.

——— (2001b): "Causes as Probability-Raisers of Processes," *Journal of Philosophy*, 98, pp. 75–92.

Scriven, Michael (1975): "Causation as Explanation," *Noûs*, 9, pp. 3–16.

Sklar, Lawrence (1977): *Space, Time, and Spacetime*. Berkeley: University of California Press.

Skyrms, Brian (1977): "Resiliency, Propensities, and Causal Necessity," *Journal of Philosophy*, 74, pp. 704–13.

Sosa, Ernest and Tooley, Michael (eds.) (1993): *Causation*. Oxford: Oxford University Press.

Suppes, Patrick (1970): *A Probabilistic Theory of Causality*. Amsterdam: North Holland.

Swain, Marshall (1978): "A Counterfactual Analysis of Event Causation," *Philosophical Studies*, 34, pp. 1–19.

Tooley, Michael (1987): *Causation: A Realist Approach*. Oxford: Oxford University Press.

——— (1990a): "Causation: Reductionism versus Realism," *Philosophy and Phenomenological Research*, 50, Supplement, pp. 215–36.

——— (1990b): "The Nature of Causation: A Singularist Account," in David Copp (ed.), *Canadian Philosophers: Celebrating Twenty Years of the CJP, Canadian Journal of Philosophy*, Supplement 16, pp. 271–322.

van Fraassen, Bas C. (1989): *Laws and Symmetry*. Oxford: Oxford University Press.

von Wright, G. H. (1973): "On the Logic and Epistemology of the Causal Relation," in P. Suppes et al. (ed.), *Logic, Methodology and Philosophy of Science IV* (Amsterdam: North-Holland).

Yablo, Stephen (2002): "De Facto Dependence," *Journal of Philosophy*, 99, pp. 130–48.